Biophysical Characterization and Identification of Small Molecule Inhibitors For the Epigenetic Reader Protein MBD2

By
Nicolás Wyhs

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
November, 2014

Abstract

Biophysical Characterization and Identification of Small Molecule Inhibitors of the Epigenetic Reader Protein MBD2

By, Nicolas Wyhs
November 2014

Significant evidence credentialing epigenetic changes as important drivers of cancer has surfaced. As a consequence, there is growing interest to find inhibitors to modulate the epigenome. Current FDA approved medications target either the DNA methylation writer protein DNMT1, or histone acetylation eraser protein family (HDAC). A longstanding interest in our lab is characterization of the 5-methylcytosine DNA methylation (meDNA) reader protein Methyl Binding Domain 2 (MBD2), a member of the MBD family and Mi-2 NuRD repression complex. Previous reports credential MBD2 as a viable cancer therapeutic target in animal and cell line models. In this thesis I describe efforts to find small molecule inhibitors of this epigenetic target.

We first characterized the biochemical parameters that can affect MBD2 binding and selectivity for methylated DNA *in vitro*, determining salt and pH substantially contribute to the affinity and selectivity of the interaction. To prepare for structure aided design we crystallized the native and selenomethionine modified methyl binding domain polypeptides of MBD2 interacting with DNA. Crystals provided data to

~2.4Å resolution. Work is underway to complete refinement and solve the MBD2-MBD structure.

With optimal conditions we tested TR-FRET and fluorescence polarization high throughput screening (HTS) assays to find inhibitors of the MBD2-meDNA interaction. Ultimately we chose TR-FRET, with Z' factors >0.7 in 384 well plates. Pilot screening with the LOPAC$^{1280™}$ library yielded an unexpected hit, NF449, which bound MBD2 in a dose dependent manner by isothermal calorimetry (ITC), with K_i in the nanomolar range.

In collaboration with Scripps Florida investigators, we applied our HTS TR-FRET assay to the NCI-MLPCN library comprising 376,276 compounds. Follow-up and counter-screens identified 18 MBD2 specific small molecules with IC_{50}'s from 1-17μM. We applied these to a TR FRET assay with multiple DNA methylation reader proteins to assess specificity. RNA expression of cancer cell lines treated with compound identified two which relieve DNA methylation mediated epigenetic promoter repression. An intercalator competition assay and ITC experiment confirmed that both bind DNA, independent of methylation status. ChIP experiments suggested that they alter chromatin to form a more open state at epigenetically repressed promoters.

In summary we will soon solve a high resolution structure of MBD2-MBD bound to DNA. We also identified at least one small molecule which binds to MBD2 and disrupts the MBD2-meDNA

interaction. Two other compounds interact with DNA and somewhat selectively disrupt DNA methylation mediated epigenetic repression.

Advisor/Mentor	William G. Nelson M.D. Ph.D.
Co-Advisor	Srinivasan Yegnasubramanian M.D. Ph.D.
Thesis Committee/Readers:	Daniel Leahy Ph.D.
	John Isaacs Ph.D.

Dedication

I would like to dedicate this work to my family:

My parents, who cared for me so lovingly for many years and made me the man I am today. And to Alec and Anna, here's for doing all you could to thwart mom and dad's work. It's been a fun ride with you guys so far, I look forward to the rest of the journey.

My F1 generation, Maya and Carina: I love every day I spend with you and can't think of anything I'd rather do than watch you grow up. You two are still my greatest experiment yet. Just give me a break in the morning and stop with the 6am wakeups! And if we have any future children sorry you weren't around to be acknowledged by name, I love you too.

And finally to my wife Sarah. Thank you for your infinite patience while obtaining this degree. It hasn't been fun with both of us in school but we finally did it! This work is truly as much yours as it is mine. I love you endlessly and look forward to a long full life of growing old with you by my side.

Preface/Acknowledgments

I arrived at the Johns Hopkins School of Medicine Department of Pharmacology after 3 years in academic and biotech research in the Boston area. After an extremely instructive rotation with Dr. Samuel Denmeade I joined the Nelson lab. I knew from the onset that I wanted a mentor who would provide me with the freedom to pursue my own questions while also acting as a safety net to catch me when I went a bit too far. Bill fit this description to a "T." I remember one of the first things he told me was that he felt his job was done when his trainees could walk into a completely foreign field, in his example ocular research at Wilmer, and ask a reasonable question that would benefit the field. Thanks to Bill's mentorship and teaching style I believe I could do that now as I depart. For never telling me my ideas or approach were ill-advised, and always being willing to discuss science when things were going well or baseball, soccer or music when they weren't, thank you! When I finally find the answer, you'll be my first call.

I was also fortunate to acquire a co-mentor when I joined the Nelson lab in Vasan. The constitutive skeptic, Vasan taught me to rigorously and objectively look at any data that we obtained in a logical and stepwise manner. As an engineer I appreciated this approach and it helped mold the way I look at asking and answering biological questions. The most important advice I ever received from him is summed up in one sentence: "The data is the data." Never try to over or under analyze it,

and if it isn't fulfilling your hypothesis find a new one. Thank you Vasan for sharing your ridiculously extensive next generation sequencing technology experience and teaching me how to troubleshoot problems and experiments.

And to my lab mates and classmates, thank you for your support over the years. Under normal circumstances I do not believe Dave, Debika, Melody or Chris and I would ever have become friends, but misery loves company and spending 5.5 years together has driven us together. Thanks for the many scientific and other discussions we've had. I also want to send a special thank you to Dr. Michael Haffner. You've spent countless hours sharing your ideas and thoughts on my experiments, life and academic progress with almost no return on your investment. You've truly been the General in the trenches with us in the lab, thank you for everything!

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Introduction

The term epigenetics derived from the word "epigenesist," first coined by Conrad Waddington in 1942. While Waddington was ahead of his time in his thinking on the topic, the coming decades have seen significant research on the levels of regulation of the genome beyond the genetic sequence (*1, 2*). This seminal leap in our thinking has drastically altered the way we think about genomic organization. While genetics provide a stable means of passing on important biological information, epigenetics affords cells a way to inherit important genomic information while remaining alterable to allow flexibility to an ever changing environment. Progress towards understanding these marks has only recently reached its full potential, as the plasticity described above creates a difficult "moving target" to try and research. While there are numerous epigenetic processes, the two best understood mechanisms are histone modifications and DNA methylation. I will consider these two mechanisms in this introduction to provide some context for the experiments performed in this thesis.

While the details surrounding the specific contributions of these epigenetic marks to cancer remain somewhat clouded, their importance for promoting cancer initiation and progression is incontrovertible(*3*). The relevance of epigenetics to cancer should not come as a surprise. Plasticity is inherently required for differentiation and development, as a single totipotent cell serves as the starting point for an entire

organism(*4*). The benefit for a cancer to hijack this pathway, to force dedifferentiation and uncontrolled growth, is very logical. The best evidence for this observation comes from monozygotic twin studies. Identical twins, by definition, share the same constitutional genetic sequence. In the extreme scenario, where genetic inheritance of defective genes is the sole cause of cancer, nearly 100% of identical twins should show concordance for developing the cancer. In reality this is not the case for almost any cancer; studies comparing concordance rates between mono- and di-zygotic twins have estimated that inherited genetic factors underlie approximately 20-50% of the pathogenesis of human cancers, but vary widely depending on the disease(*5-10*). The remaining risk is likely due to environmental factors, which can influence epigenetic processes. With such a large fraction of total risk potentially comprised of epigenetic changes it is unsurprising that this field has seen renewed interest in recent reports.

Histone Modifications as an Epigenetic Control Mechanism

The first description of histone modifications appeared in 1964 in a paper defining the presence of methylated lysines on histone tails(*11*). A series of articles describing acetylation and phosphorylation marks, also on lysine, quickly followed (*12, 13*). Since their announcement, the importance of histone changes for normal and malignant processes has

been shown repeatedly(*3, 14, 15*). Despite significant progress new modifications are frequently reported.

The most basic unit of chromatin organization is the nucleosome, a complex comprised of four paired histone subunits (H1, H2A, H3 and H4)(*16*). 147 base pairs of DNA wrap around each nucleosome, providing the cell with a means of both compaction and transcription control(*16*). By marking the histone tails, which serve as a sort of flag, regions of the genome are labeled for dense compaction with limited access and expression (heterochromatin) or loose compaction for easy access of transcription factors (euchromatin)(*17*). Cancers often corrupt writer and eraser proteins which alters these marks, forcing the genomic landscape near tumor suppressors to form heterochromatin establish euchromatin at the regulatory regions of pro-oncogenic genes(*18*). These changes drive un-policed proliferation and dedifferentiation, both hallmarks of cancer(*18*).

A significant effort towards finding therapies to prevent this commandeering, or reverse the inappropriate histone changes has occurred(*19*). To date only 2 histone directed medications have gained FDA approval, and both target the eraser class of proteins known as histone deacetylases (HDAC)(*20, 21*). While these drugs clearly benefit patients with some human malignancies such as cutaneous T-cell lymphomas, a significant amount of work remains to find better and

more widely applicable small molecules. HDAC inhibitors and their use are described in more detail in chapters 3 and 4.

DNA Methylation Alterations

The second key epigenetic mechanisms to control gene expression is DNA methylation(*15*). The first description of DNA methylation in eukaryotic cells appeared just prior to the discovery of histone modifications in the 1950's(*1,22,23*). This epigenetic mark occurs at the 5-position of cytosines (5meC), located in a CpG dinucleotide pair(*1*). The application of the mark is the responsibility of a family of DNA writer proteins referred to as the DNA methyltransferases (see below) and requires both ATP for energy and *S*-adenosyl-methionine (SAM) as a methyl donor(*1*). Increased frequency of the CpG sequence is often found at gene promoters and regulatory sequences(*24*). These enriched areas are referred to as CpG islands, and their dense methylation leads to heterochromatin through the actions of methylation reader proteins (see below)(*24*).

Histone and DNA methylation modifications are tightly linked. One mechanism of epigenetic inheritance from parent to daughter cells suggests that 5meC modifications are more "permanent," because they are found at regions where heterochromatin is maintained for a longer period of time (*25*). In contrast, histone modifications are considered

malleable as they can be quickly removed by known enzymes(*26*). Recent models question this assumption, though, suggesting DNA methylation may also exhibit relatively rapid cycling under specific cell conditions (*27*). Ultimately we believe the interplay between three classes of proteins: readers, writers, and erasers, governs the lifecycle of DNA methylation and histone marks. Since the focus of this thesis is DNA methylation changes we will describe the readers, writers and erasers for this modification below, and leave the discussion of histone modifications for another venue.

Writers of 5-Methylcytosine on DNA

DNA methylation writers are collectively known as the DNA methyltransferases (DNMTs). There are three established members, all capable of catalytically methylating DNA: DNMT1, DNMT3a and DNMT3b(*28*). Two other proteins share homology but lack enzymatic function: DNMT2, shown to be essential for aspartate tRNA methylation (*29*), and DNMT3L, an essential DNA methylation mediator discussed later.

DNMT1 is referred to as the maintenance DNA methyltransferase. This protein associates with the replication fork and the histone protein modifier G9A and assists with maintaining the pattern of methylation to daughter stand DNA(*30*). During semi-conservative DNA replication,

newly replicated DNA is initially hemi-methylated immediately following replication, where the original strand is methylated but the newly created strand is not. DNMT1 converts the hemi-methylated CpG dinucleotide to fully methylated, thus maintaining the epigenetic information(*30*). Detailed crystallography research elucidated the enzymatic mechanism of DNMT1(*31*): The CXXC-BAH1 linker region is responsible for the preference of DNMT1 for hemi-methylated(*31*). When bound to an unmethylated oligonucleotide, the linker is positioned at the 5-carbon of the unmethylated cytosine in the active site, which sterically hinders enzymatic function (*31, 32*). If the oligonucleotide is hemi-methylated, the linker cannot insert into the active site and the cytosine is optimally positioned for methylation.

Due to the essential function of DNMT1, a number of studies have verified its benefit as a therapeutic target. Knockout of DNMT1 in mouse embryonic stem cells shows a 3 fold reduction in methylation levels, While a $Dnmt1^{-/-}$ genotype in a mouse is embryonic lethal (*33*), mice with hypomorphic DNMT1 alleles display protection against intestinal tumorigenesis, without showing severe developmental abnormalities (*33*). This suggests that the enzyme is essential for development, but that some residual activity may be sufficient to sustain these developmental processes. Murine embryonic stem cells lacking DNMT1 contain a large number of hemi-methylated genomic regions (*34*). This suggests that even in the presence of other DNMTs, DNMT1 is required to convert

hemi-methylated DNA into a fully methylated state(*35*). From a targeting perspective, inhibiting DNMT1 can be effective for inhibition of cancer cell growth, but also can have toxicity to cells that proliferate and require maintenance of DNA methylation.

DNMT3a and DNMT3b are responsible for *de novo* DNA methylation, actively converting unmethylated DNA into hemi-methylated DNA *in vivo*(*28*). Each is expressed in a number of alternatively spliced constructs, but the role of each alternate construct remains to be fully understood (*36, 37*). Despite their difference in size, each DNMT3 writer contains a homologous C-terminal methyltransferase domain(*36*). Attempts to create DNMT3 knockout mice showed similar results as DNMT1: the Dnmt3a$^{-/-}$ genotype is lethal during the early stages of postnatal development (*28*), while Dnmt3b$^{-/-}$ mice are embryonic lethal, at a similar stage as DNMT1$^{-/-}$ knockouts (*28*). In both cases, significantly lower levels of genomic methylation are seen (*28*). Dnmt3b mutations have also been associated with a known human syndrome called immunodeficiency, centromeric instability, facial anomalies or ICF (*28*). The severe phenotypes accompanying these diseases suggest that targeting DNMT3's could lead to significant toxicity. As discussed later, there are indeed dose limiting toxicities when using current DNMT inhibitors.

DNA methyltransferase 3-like protein (DNMT3L) lacks the catalytic domain of other DNMT3 family members, and is structurally distinct

from DNMT3a and 3b despite sharing significant sequence homology(*38*). DNMT3L interacts directly with DNMT3a and unmethylated histone H3K4 tails via its ATRX-DNMT3-DNMT3L (ADD) domain, forming a tetramer of two DNMT3A molecules each bordered by one DNMT3L (*38, 39*). The current functional model proposes that DNMT3L binds unmodified H3K4 and recruits two DNMT3A proteins which are positioned approximately one helical turn apart on DNA (*40, 41*). Furthermore, this reported 3L-3a-3a-3L tetramer was observed in long nucleoprotein filaments when combined with DNA, suggesting a mechanism for rapid *de novo* methylation(*41, 42*). This mechanism is supported by biophysical studies which show combining DNMT3L and DNMT3 enzymes improves *de novo* methylation activity up to 20 fold over the enzymes alone *in vitro*(*43*). This observation supports previous studies that DNMT3's are most important during differentiation and rapid cell growth, events that occur during development. (*44*).

DNMT3L is highly expressed throughout embryogenesis, where its expression mirrors that of DMNT3a/b and accompanies the rapid increase in genomic methylation observed at key steps during development(*45*). Dnmt3L$^{-/-}$ female mice have no specific phenotype, but their progeny are almost all lost during gestation, even when crossed with a wild type male(*46*). Male Dnmt3L$^{-/-}$ mice are sterile but otherwise asymptomatic(*46*). Interspecies conservation suggests that only animals with maternal imprinting patterns have a strongly conserved DNMT3L

sequence, suggesting that it is involved in maintaining imprinting patterns(*40*). While inhibiting DNMT3L would most likely not halt *de novo* methylation, it could lead to inefficient methyltransferase activity. Presumably, this would alter genomic regions that are difficult to methylate most effectively.

DNA Readers

5-methylcystosine marks are recognized by three major families: SRA domain-containing proteins, zinc finger and BTB-containing (ZBTB) proteins, and methyl CpG binding domain (MBD)-containing proteins (*47*). The MBD family contains seven members: MBD1, MBD2, MBD3, MBD4, MBD5, MBD6 and MeCP2(*48*). The existence of methylated-DNA binding proteins was first published in 1989, and followed by identification of the first family member, methyl CpG binding protein 2 (MeCP2), in 1992 (*49, 50*). All family members contain a conserved methyl binding domain; despite this, it is believed that three members, MBD3, MBD5 and MBD6, are not involved in binding methylated DNA, but may be associated with various repression complexes(*51*).

After identification, MeCP2 was shown to interact with Sin3A and indirectly with HDAC1 and HDAC2 through other complex members to coordinate a repressed chromatin state in the vicinity of the DNA methylation marks(*52*). This landmark discovery was the first to describe

a DNA methylation “reader” mechanism connecting DNA methylation and gene repression. MeCP2 is minimally expressed in development but is prevalent as cells differentiate, with expression changing concomitant with rising global methylation levels(*53*). This expression pattern is a common theme among proteins associated with DNA methylation, including DNMTs and other MBD family members. The rise of 5-methylcytosine marks in differentiated cells requires a critical amount of epigenetic machinery to functionally suppress gene expression(*54*).

Mutations or loss of MeCP2 in human males is lethal, and in females confers a particularly severe neurological disease known as Rett Syndrome(*55*). This syndrome typically presents around age 2, and leads to significant muscle ataxia and atrophy with growth and learning delays(*55*). Mice with Mecp2 null mutations appear to develop normally but after 6 weeks of age begin to display significant neurological deficiencies and retarded development, similar to Rett patients (*56*). This phenotype is not salvageable with increased MeCP1 levels, and is recapitulated in a conditional knockout of neural tissue (*56*). The severe side effects associated with MeCP2 disruption may have contributed to the very limited work done to find relevant inhibitors.

MBD1 was identified as the second methyl binding domain family member and was initially called Protein Containing MBD 1 (PCM1) (*57*). Its detection came using a new technology at the time, online homology searching using EST tags, and was subsequently purified and shown to

cause a super shift of the MeCP1 complex when bound by an antibody (*57*). Despite initial experiments suggesting MBD1 as a member of the MeCP1 complex, it is now established as associating with SETDB1, Suv39h1 and HP1 to cause chromatin alterations and gene silencing independent of this complex(*58*).

In addition to MBD and transcription repression domains, MBD1 contains three CXXC zinc finger domains that are thought to bind specific DNA sequences(*59*). A brief BLAST search of these domains showed that two have sequences with slight alterations to the canonical CXXC domains found in proteins like MLL and DNMT1, but the third domain is almost identical and can bind unmethylated DNA(*59*). Further work suggested that it is also responsible for decreasing transcription at unmethylated promoter regions, an attribute no other MBD family member boasts(*60*). MBD1 may therefore repress expression using both epigenetic and non-epigenetic reader mechanisms. Practically, however, it is unclear if MBD1 uses both domains together or at all. $Mbd1^{-/-}$ mice are fertile and developmentally normal, suggesting that it is not absolutely required for normal cellular function and could have limited side effect toxicity if targeted for therapy.

MBD2 is another DNA methylation reader protein critical for epigenetic repression(*51*). It was discovered in 1998 by Bird et al through homology searching and established as the third MBD family member(*51*). This ~50kDa protein is a member of the MeCP1 complex,

also referred to as the Mi-2/NuRD complex, where it is responsible for reading DNA methylation and transmitting its repressive signal to histone modifying proteins to form heterochromatin(*61, 62*). *In vitro* binding data suggests that MBD2 in isolation is capable of binding a single fully methylated CpG, suggesting that its binding specificity is inferred from other proteins in the complex (*63*). MBD2 preferably binds fully methylated CpG's over hemi-methylated dinucleotides and recruits Mi-2/NuRD complex and associated histone deacetylases HDAC1 and HDAC2(*61*). It is thought that the signaling of both MeCP2 and MBD2 reader proteins occurs through ordering of the MBD protein structure upon identification of a methylated substrate((*64*), and Wyhs et al unpublished data). In solution, reader proteins sample multiple conformations and are relatively unorganized. Upon recognition of a methylated cytosine substrate, they assume a more ordered and rigid position, which allows MeCP1 complex members to organize in a conformation that promotes optimal HDAC activity for subsequent histone deacetylation, heterochromatin formation and gene silencing(*64*). Similar to MeCP2 and MBD1, MBD2 is not expressed in embryonic stem cells, but can be found in increasing abundance as the embryo differentiates during development(*65, 66*). $Mbd2^{-/-}$ mice are healthy and fertile, and no known human syndrome exists for MBD2 mutations(*66*). This suggests that targeting MBD2 for cancer therapy may have an attractive toxicity profile (*66, 67*). $Apc^{min/+}$ mice carrying homozygous

disruption of MBD2 alleles develop significantly fewer tumors and have increased survival compared to $Apc^{min/+}$ mice with wild type MBD2 alleles(*67*). This suggests that targeting MBD2 with small molecules could be beneficial in the setting of familial adenomatous polyposis, or even for prevention or treatment of other neoplasms.

Unlike MeCP2, which is a relatively small complex with only a few proteins, MeCP1 is quite large with at least 18 known proteins encompassing about 1MDa of weight (*68*). Optimal binding conditions for MeCP1 exist when multiple CpG's are methylated in the same area, such as a CpG island at a promoter region (*49*). This is different from MeCP2, which is able to identify a single methylated CpG dinucleotide as effectively isolated as it can with its complex members (*69*).

Though it contains a methyl CpG binding domain, MBD3 is unable to interact with methylated substrates due to a substitution of phenylalanine for tyrosine at the 23rd residue(*70, 71*). X-ray crystallography structures of MeCP2 and NMR structures of MBD1 and MBD2 suggest that this residue provides a critical water-mediated contact necessary for identification of methyl-cytosine(*72, 73*). The loss of a hydroxyl group at this residue suggested that MBD3 may instead bind hydroxymethylcytosine, but this theory has not been substantiated(*74*). Although MBD3 does not bind methylated DNA, $Mbd3^{-/-}$ mice are embryonic lethal, suggesting it has an as yet unknown crucial role in development(*66*).

The endonuclease MBD4 protects the genome against spontaneous methyl cytosine deamination, which converts the base into thymidine(*75*). Spontaneous hydrolytic deamination occurs at the rate of approximately $5.8x10^{-13}$ s^{-1}, or if one assumes 1% of the genome is 5-methylcytosine, about 2-3 deaminations per day per cell(*76*). If left unchecked, a cell would lose not only its methylation status but be unable to tell which thymidine bases were once 5-methylcytosine. Such substitution mutations could be disastrous if left unchecked. MBD4 preferentially binds to T-G mispairs and utilizes a c-terminal thymine DNA glycosylase domain to remove the inappropriate thymine base pair; this allows normal cell DNA repair machinery to replace it with a cytosine via its interactions with MLH1(*77, 78*). DNMT1 can then re-methylate the new cytosine to maintain epigenetic fidelity. *In vivo* modeling with $Mbd4^{-/-}$ mice have a threefold increase in the rate of C to T transversions, but show only mild phenotypic changes during development(*79*). When $Mbd4^{-/-}$ mice are crossed to the spontaneous intestinal tumor model $Apc^{Min/+}$ offspring actually show an increased tumor burden and a higher number of mutations per gastrointestinal cell for both tumor and normal tissue(*79*). While MBD4 does not seem to interact with methylated DNA like MBD2 and MeCP2, it has a clear role in maintaining DNA methylation and DNA sequence fidelity.

Similar to MBD3, MBD5 and MBD6, the two newest members of the MBD family, do not interact with methylated DNA (*48*). The MBD for

these proteins contains an intron dividing it in half(*48*). This leads to a nine residue deletion upstream from a beta sheet that is involved with methylcytosine binding in other family members. This intronic insertion, along with two residue alterations in another critical region of the methyl CpG binding domain, likely contribute to the inability of MBD5 and MBD6 to bind methylated DNA *in vitro* and *in vivo* (*48, 80*). There are no mouse knockout studies of MBD5 to date, but its mutations are associated with human neurological diseases, including seizures and autism (*81*). Similarly, MBD6 directly interacts with ATXN1L, a protein that has been associated with the neurological disorder spinocerebellar ataxia type 1(*48*). A good deal of work remains to decipher the purpose and mechanism of MBD5 and MBD6.

Separate from the MBD family there is another well-studied DNA methylation mediator: Ubiquitin-like containing PHD and RING finger domains 1 (UHRF1)(*82*). It is a multi-domain protein that interacts with both histone tails and methylated DNA. It contains a Tandom Tudor (TT) domain, shown to interact with the histone H3 in various states of modification (*83*), a Set and RING-finger Associated (SRA) domain that has been crystalized with a hemi-methylated DNA oligonucleotide (*83*), and a Really Interesting New Gene (RING) domain that acts as an E3 ubiquitin ligase, having been shown to ubiquitylate DNMT1 and, interestingly, histone H3K23, which is thought to be necessary to complete DNA methylation maintenance (*82, 84*). UHRF1 co-localizes

with the replication fork complex, binding nascent hemimethylated DNA arising during semiconservative replication(*82, 85*). The affinity of UHRF1 for hemi-methylated DNA is thought to trigger H3K23 ubiquitination, which then increases the affinity for DNMT1 to complex with UHRF1, recruiting the maintenance methylase to fully methylate the DNA (*84*). Uhrf1$^{-/-}$ mice are embryonic lethal at gestation day 9.5.

DNA Methylation Erasers

Significant research into locating the erasers of DNA methylation has taken place over the last 5-10 years. Despite a plethora of work little direct evidence has surfaced. The leading theory, supported by a number of reports, is conversion of methyl groups to hydroxymethylcytosine, formylcytosine, and carboxycytosine by the TET family enzymes, and subsequent removal of these oxidized forms of methylcytosine by the nucleotide excision repair machinery(*86*). An excellent review on this subject by Ooi et al is available for readers who are interested in this subject(*87*).

Conclusions

Despite evidence that DNA methylation reader proteins could act as effective therapeutic targets no substantial effort has been reported. For a number of years our lab has focused on characterization and

targeting of the MBD2 reader protein. Previous students focused on cell based screens with some success. While evaluation of these compounds continues to yield fruitful avenues, I chose to approach this question from a biochemical standpoint. During the course of my thesis work, described here, I design and implement a high throughput screen to discover small molecule inhibitors that prevent MBD2 from interacting with methylated DNA. Ultimately I identified two classes of inhibitors, one which binds MBD2 and another which binds DNA but causes somewhat selective chromatin changes only at methylated promoters. In the final chapter I spend some time discussing use of the MBD2 protein as a tool to precipitate genomic methylated DNA coupled with next generation sequencing to analyze the prostate cancer methylome. The following chapters present this body of work in some detail.

Chapter 2

Biophysical Characterization and Structural Analysis of the Methyl Binding Domain of MBD2

2.1 Abstract

The biophysics behind the interaction of MBD family members, including MBD2, and methylated DNA is not completely understood. Recent advances in the NMR solution structures of the methyl binding domains for MBD1, MBD2, and MeCP2 as well as the crystal structure of MeCP2-MBD and MBD4-MBD help our understanding of this complicated binding event. These studies support the hypothesis that organized water molecules, through hydrogen bonding interactions with as few as 2-3 residues, ultimately mediate the identification of the methylated cytosine on DNA. We investigated the binding specificity and sensitivity of the MBD2-MBD protein and determine a high resolution structure with x-ray crystallography.

Comparison of the binding sensitivity and specificity of eukaryotic (insect) versus bacterial recombinant MBD2-MBD protein indicated that the insect cell version was slightly more sensitive and selective. Using a fluorescence polarization based binding assay we tested numerous buffer conditions *in vitro* and found that altering salt concentrations above 150mM decreased sensitivity, while going below 100mM decreased specificity. Similarly, lowering the pH led to loss of specificity, while higher pH led to loss of sensitivity. We combined these results to discern an optimal buffer for designing a high throughout assay for identification of small molecule inhibitors of MBD2 (chapter 3-4), and for utilizing this protein as a tool for methylated DNA isolation (chapter 5).

We continued our work with MBD2-MBD and successfully formed stable crystals in complex with methylated DNA in 1.4M citrate pH 6.5. X-ray diffraction data were collected to ~2.4Å resolution. Molecular replacement failed to provide a suitable solution for phasing, and numerous experimental techniques were employed. Two selenomethionine proteins were successfully crystallized, MBD2-MBD-R195M and MBD2-MBD-L193M, and SAD data were collected for each. 3 selenomethionine sites were found with the L193M construct and overlay of the MeCP2-MBD solution with preliminary electron density maps suggests that protein and DNA are present. In collaboration with the lab of Dr. Dan Leahy, refinement and modeling is ongoing towards finding a solution for x-ray structure of MBD2-MBD.

2.2 Introduction

As discussed in the introduction MBD2 was discovered by Bird et al in 1998 through homology searches after characterization of MeCP2 and MBD1(*51*). This initial report established the basic framework behind the biophysics of MBD2 and the MBD protein family and how they physically interact with methylated DNA. The first biophysics experiment which demonstrated MBD2 directly interacts with meDNA was a gel shift assay in the discovery report: in the presence of protein, 5-methylcytosine labeled DNA shifted upward to a higher molecular weight (*51*). This was corroborated by immunofluorescence (IF) showing MBD2 labeled with green fluorescent protein (GFP) co-localizing to major satellite repeat regions known to be highly methylated in embryonic stem cells (*51*). The result led to a number of avenues of research including structure, the kinetics of binding, and specificity of binding related to DNA. Since the late 1990's a number of publications answer some of these questions and we will discuss them here.

A bit of controversy surrounds the details for MBD2 binding sequence specificity and the optimal number of methylated CpG dinucleotides required. The initial discovery of the MeCP1 complex, which includes MBD2, required 12 CpG pairs for adequate binding (*88*). MeCP2 on the other hand, required only one methylated dinucleotide pair for interaction with DNA (*88*). Work in our lab indicates that when MBD2 is isolated from the MeCP1 complex, however, it also recognizes

and binds singly methylated CpG sites (*63*). A number of studies support our finding indirectly and corroborate the theory that the specificity of MBD2-MBD protein binding is predicated only on the basis of 5-methyl CpG content, and not dependent on DNA sequence. In support of this idea Jørgensen et al demonstrated that by engineering a multi MBD domain protein they could drastically improve the sensitivity of protein binding to methylated CpG's without having to mutate the primary protein sequence or compromising specificity (*89*). Whole genome ChIP experiments also support this finding with the binding sites of the MBD domain containing family members MBD1, MBD2 and MeCP2 enriched at highly methylated regions more so than at any other particular sequence or modification (*90*). Interestingly, however, the sites occupied between the MBD's did not overlap perfectly (*90*). Other reports also demonstrate that the different MBD family members cannot replace each other at particular genomic loci under all circumstances (*91*). A review of clinical evidence supports this observation: females who have mutated MeCP2 genes suffer from a significant neurological disorder known as Rett Syndrome despite having normal MBD1 and MBD2 proteins(*92*). Taken together this evidence strongly suggests that there must be a yet described specificity provided by either the other members of the various repression complexes or the MBD protein itself, and that a stochastic model of binding cannot alone explain all of the observations presented in the literature.

A pair of reports published during the 2000's provide possible answers to this question. In one, the authors demonstrate that while MBD2 seems to have no sequence specificity, MeCP2 preferred A-T rich sequences adjacent to a methylated cytosine (*91*). The second reveals MBD1 may attain some sequence preference by utilizing a second DNA binding domain contained within its protein sequence, a zinc finger based CXXC3 domain. This extra CXXC3 domain could help identify particular DNA sequences, enhancing its specificity for particular loci in the genome(*93*). Despite these provocative reports, a lot of work remains to confirm the results for both MeCP2 and MBD1 and elicit how, if at all, MBD2 is targeted to particular locations on the genome.

Since the discovery of the MBD family a number of structural studies have been reported. The first work describing the mechanism by which MBD family members identify methylated cytosines was published in 1999. Through careful foot printing experiments using restriction enzymes and nucleosomal DNA, Chandler et al established that MeCP2 utilizes its MBD domain to identify methylated cytosines through the major groove (*94, 95*). They also suggested that the C-terminus makes contact with both strands of non-nucleosome bound DNA, providing a plausible answer to why the protein only binds symmetrically methylated DNA and not single stranded DNA or RNA(*95*). Two reports later in the same year solved the NMR solution structures of the MeCP2 and MBD1 methyl-CpG-binding domains(*96, 97*). They corroborate previous findings

that the proteins likely identify the modified DNA through its major groove. Both domains consist of a cluster of 4 beta sheets and one alpha helix. The alpha helix and DNA facing portion of the beta sheets are scattered with lysines and arginines, consistent with previous reports describing the interaction between the MBD domains and DNA to be independent of base sequence, but rather driven by charge interactions between the phosphate backbone and protein). The residues on the last pair of the beta sheets are predicted to form a hydrophobic pocket consisting of a tyrosine and isoleucine (or glutamine in MBD1) along with an alanine that could mediate van-der-Waal's interactions with the methylated cytosine located nearby. Unlike many other transcription factors that bind to palindromic DNA sequences, MeCP2 and MBD1 do not have noticeable symmetry in their secondary or tertiary structures (*96, 97*). Furthermore, neither protein binds as a dimer, further clouding the question of how these domains maintain specificity and sensitivity alone. Regardless, these initial reports provided many insights into the physics of this interaction and helped guide future experiments.

It was almost a decade later, in 2008, that the first crystal structure of an MBD was reported(*72*). The MeCP2 methyl binding domain was solved at 2.5Å resolution complexed with a 20bp symmetrically methylated CpG oligo sequence from the BDNF gene promoter. The crystal structure very accurately overlaid with the NMR model, confirming previous descriptions. Surprisingly, only 3 residues

directly interact with DNA bases: asp121, arg111, and arg133(*72*). Asp121, and tyr123 identify the methylated cytosine, the previous via a direct hydrogen bond, and the later through water mediated hydrogen bond. Mutations of tyr123 to phenylalanine in MeCP2, or the homologous tyrosine in MBD1 to Y34F, show significantly lower methylated cytosine affinity by gel shift assay (*72*). Interestingly, in wild type (wt) MBD3, previously established as not binding methylated cytosine labeled DNA, this tyrosine was already a phenylalanine (*70, 71*). MBD3-F34Y mutations regain the ability to bind 5-methylated CpG dinucleotides, confirming the importance of this residue and water in mediating this protein DNA contact(*71*). In MeCP2 R111 and R133 make hydrogen bond contact with the symmetric guanine bases associated with the methylated cytosines, exposing the mechanism behind the dinucleotide symmetry required for this binding interaction(*72*).

Despite solving a higher resolution structure for MeCP2-MBD we are still unable to provide any context as to why there could be sequence specificity outside of the methylated CpG dinucleotide. This strengthens the argument that the methyl binding domain proteins themselves, in particular MBD2 and MeCP2, may not prefer a specific sequence. Another remarkable experiment from this work revealed that the MeCP2-MBD T158M mutation, found in patients suffering Rett Syndrome, significantly lowered the affinity of MeCP2 protein to DNA in a gel shift assay. This result lends itself as a possible mechanism for the dramatic

phenotype seen in these patients, and validates the MBD domain as a possible target for small molecule disruption to interrupt protein function.

The only other MBD x-ray crystallography structure solved to date is that of MBD4-MBD(*98*). As discussed earlier, this protein is critical for repairing G/T mismatches that occur as a consequence of deamination of a 5 methylated cytosine(*77*). The structure was solved complexed with a symmetrically methylated CpG dinucleotide as well as a meCpG/TpG mismatched oligo sequence. The solution shows that the MBD4-MBD binding pocket is extremely versatile, utilizing a number of water molecules to make critical contacts between the DNA and protein instead of direct hydrogen bonding. This allows it to conform to the many substrates it interacts with (C/G, T/G, meC/G etc.). Surprisingly this protein also binds to hydroxymethylcytosine –guanine dinucleotides, a result that should be investigated further.

Structural work on MBD2 has not been as fruitful as for MeCP2. A solution structure of the chicken MBD2 methyl binding domain bound to a methylated *p-globin* DNA promoter was described in 2011 to about 6Å resolution(*99*). Overlay of the MBDs between the MBD2 and MBD1 solution and MeCP2 crystal structure support their high level of homology, with only minor differences noted. For example the critical tyrosine 34, required for methylated cytosine binding in MBD1 (Y123 in MeCP2), is slightly closer in MBD2, potentially allowing it to make a

direct hydrogen bond to the methyl group(*99*). MBD1 is also slightly rotated, allowing it to contact the DNA phosphate backbone between its β1 and β2 loops, where MBD2 will likely not be able to reach(*99*). MeCP2's and MBD2's methyl binding domains are virtually identical in their binding pocket. The only predicted, minor, alterations are to the alpha helix, with MeCP2's being longer by 4 residues. The authors go on to state that for this particular promoter region MBD2 seemed to bind in one particular orientation on the positive strand(*99*). While they did not confirm their results with experimental data, it is intriguing to think this could imply strand specificity for MBD2 MBD.

Despite the significant amount of progress made in the past decade in understanding the biophysical interaction of MBDs and meDNA, a number of questions still loom. In this work I focused on MBD2 and asked how its biophysical properties change under different conditions. In doing so we decipher the optimal conditions for targeting MBD2 and using it as a tool (discussed in more detail in chapters 3, and 5). Interestingly the results also shed some light on ways to alter the protein's specificity or sensitivity for meDNA. To better understand the protein, and also help us with possible structure aided design, see chapter 3 and 4, we set out and successfully collected x-ray diffraction data on an MBD2-MBD selenomethionine and wild type crystals to approximately 2.4Å. While this work is still on going, we're optimistic to have a solution to the underlying structure very soon.

2.3 Results

Comparison of Recombinant MBD2-MBD from Bacterial and Insect Cells

As an eager first year graduate student I entered the Nelson lab and was quickly taught how to purify recombinant MBD2-MBD protein. We initially generated recombinant purified MBD2-MBD using a baculoviral/insect cell system. We expressed protein in SF9 insect cells through a baculoviral vector containing residues 145 to 213 of MBD2. Unfortunately this system left us with a relatively low yield of protein (µg per liter final product) and so I opted to move towards a bacterial expression system to improve yield. After successfully expressing MBD2-MBD in BL21-DE3 cells, we purified recombinant protein as described in the methods with yields of 5-10mg/L. As a eukaryotic cell, the SF9 line potentially adds post translational modifications to the expressed protein. Previous work in our lab, the literature, and via queries with online prediction tools all suggest that MBD2 is phosphorylated within the MBD domain (Reichert et al unpublished, (*100*)(*101*)). Only one post-translation modification to date is well investigated though: a methylation mark on R61 in the GR repeat domain(*102*). Tan et al demonstrate that MBD2 interacts with two arginine methyltransferases, PRMT1 and PRMT5, and that they methylate the R61 of MBD2 and lower it's affinity for meDNA by gel shift assay. Review of the data shows that

while this observation holds, it is not fully corroborated by other reports and the difference in binding affinity is not strong.

To investigate whether recombinant MBD2-MBD purified from a eukaryotic host (potentially modified post-translationally) vs a prokaryotic host (not post-translationally modified) could alter binding to meDNA we performed a gel shift assay using MBD2-MBD made in insect cells, as well as our new bacterial system (Figure 2.1). We also wanted to determine if we could manipulate the only cysteine contained with our construct (C9) for other applications so N-ethylmaleimide treated protein was included. Interestingly, at the conditions investigated, the insect cell protein appeared to have similar, but higher affinity, compared to the bacterial protein. This supports previous reports that post translational modifications could impact MBD2's ability to bind methylated cytosine(*102*). N-ethylmaleimide treatment significantly disrupted protein binding to methylated DNA, suggesting that use of this cysteine for labeling may not be ideal.

MBD2-MBD Binding Under Different Physiological Conditions

In order to utilize MBD2-MBD as a tool for enriching methylated DNA (see chapter 5) and also to find optimal conditions for *in vitro* screening of small molecule inhibitors (chapter 3) we tested the ability of bacterially made MBD2-MBD to bind a single symmetrically methylated

CpG DNA under different conditions *in vitro*. We monitored the interaction using a fluorescence polarization assay with a fluorescein (FAM) labeled oligo. The first variable tested was salt (NaCl) concentration. Lower concentrations typically made the protein more

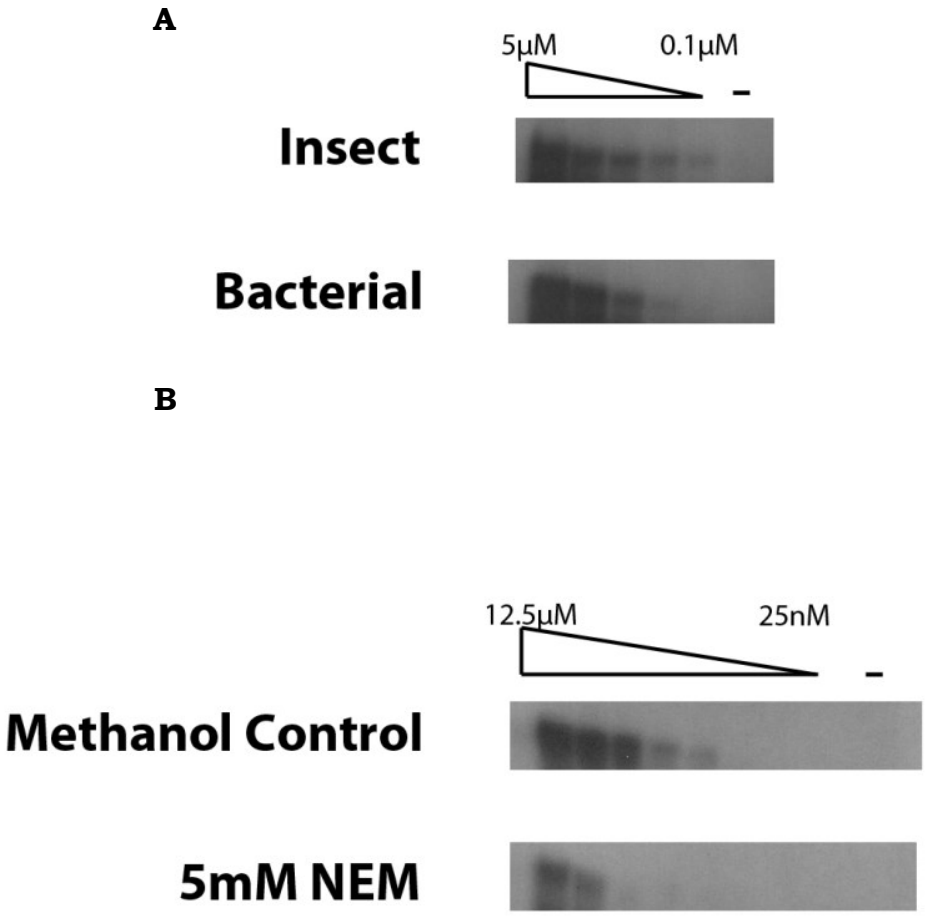


Figure 2.1: Gel shift of recombinant MBD2-MBD from insect and bacterial cells with meDNA

(A) MBD2-MBD made in either insect or bacterial cells binding to a symmetrically methylated hairpin oligo in an increasing dose response. **(B)** Increasing concentrations of bacterial MBD2-MBD treated with cysteine modifying agent NEM.

promiscuous, lowering its specificity (Figure 2.2A). Higher salt lowered the overall sensitivity of the protein for DNA (Figure 2.2A), as observed by lower total binding. The Z' statistic determined the significance of these results. Comparing protein binding to unmethylated and methylated substrates 100-150mM NaCl was the most significant (Table 2.1). These results are consistent with what would be expected: as the ionic strength of the solution lowers the positively charged protein will seek out other charges, with the most abundant being the negatively charged phosphate backbone, leading to non-specific binding. The non-specific binding is clearly seen in figure 2.2A upper left panel as the low salt protein-meDNA interaction never fully saturates. We confirmed that the type of salt is not as important as the overall ionic strength by observing no differences in sensitivity or specificity when maintaining an optimal concentration of 125mM salt but substituting increasing amounts of KCl in place of NaCl (Figure 2.3).

The next variable analyzed was pH, also performed using our fluorescence polarization assay. MBD2 was exposed to different buffers at pH levels ranging from 4.5-10.5 (Figure 2.4). Interestingly pH followed a similar trend to that observed with salt concentration. Increasing pH leads to a loss of sensitivity (Figure 2.4A) whereas decreasing pH lead to a loss of specificity (Figure 2.4B). Unsurprisingly the optimal balance between these effects is achieved at the physiologically relevant pH of approximately 7.5.

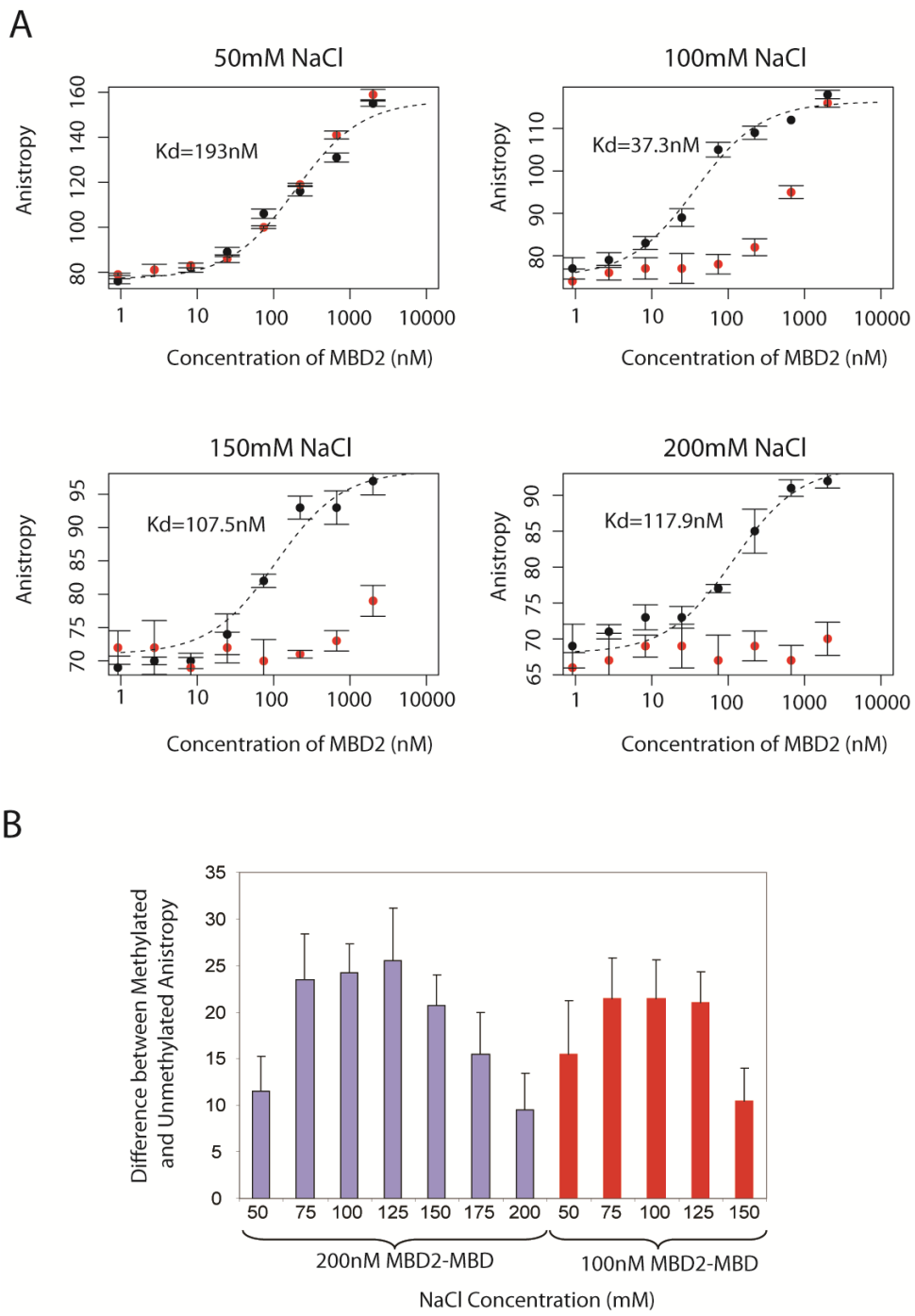


Figure 2.2: MBD2-MBD and Methylated DNA binding under different salt concentrations

(A) MBD2-MBD binding to singly symmetrically methylated hairpin DNA oligos measured by fluorescence polarization assay. Experiments were run at 50mM, 100mM, 150mM or 200mM NaCl concentrations. **(B)** Difference between methylated and unmethylated DNA binding *in vitro* at various salt concentrations and 200nM or 100nM MBD2-MBD by fluorescence polarization.

NaCl Concentration (mM)

MBD2-MBD Concentration (nM)	50mM	75mM	100mM	125mM	150mM	175mM	200mM
1	-0.73	-	-4.40	-	-2.82	-	-4.14
3	-	-	-2.46	-	-9.39	-	-2.92
8	-8.00	-	-1.02	-	-3.83	-	-1.67
25	-2.73	-	-0.32	-	-5.04	-	-3.12
74	-0.50	-	0.55	-	-0.08	-	-0.31
100	-0.11	0.40	0.43	0.53	-	-	-
200	0.01	0.37	0.62	0.34	0.53	0.14	-0.25
222	-1.18	-	0.60	-	0.68	-	0.04
667	-0.12	-	0.73	-	0.41	-	0.60
2000	-1.60	-	-2.00	-	0.27	-	0.56

Table 2.1: Z' Statistic for MBD2-MBD and meDNA at various salt and protein concentrations

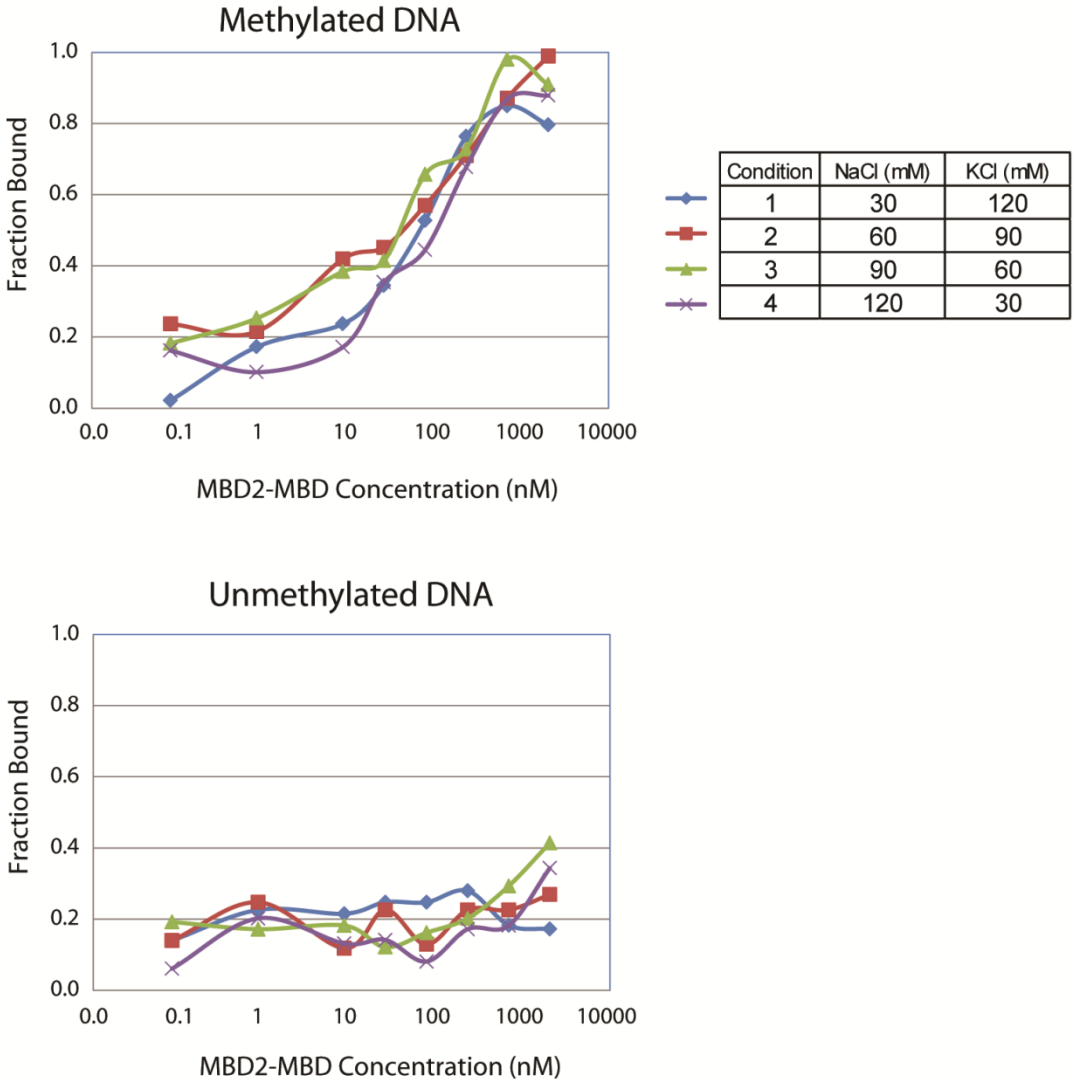


Figure 2.3: MBD2-MBD binding to methylated DNA at constant ionic strength

MBD2-MBD binding to methylated DNA (upper) and unmethylated DNA (lower) at constant ionic strength (125mM) but different NaCl and KCl concentrations as measured by fluorescence polarization.

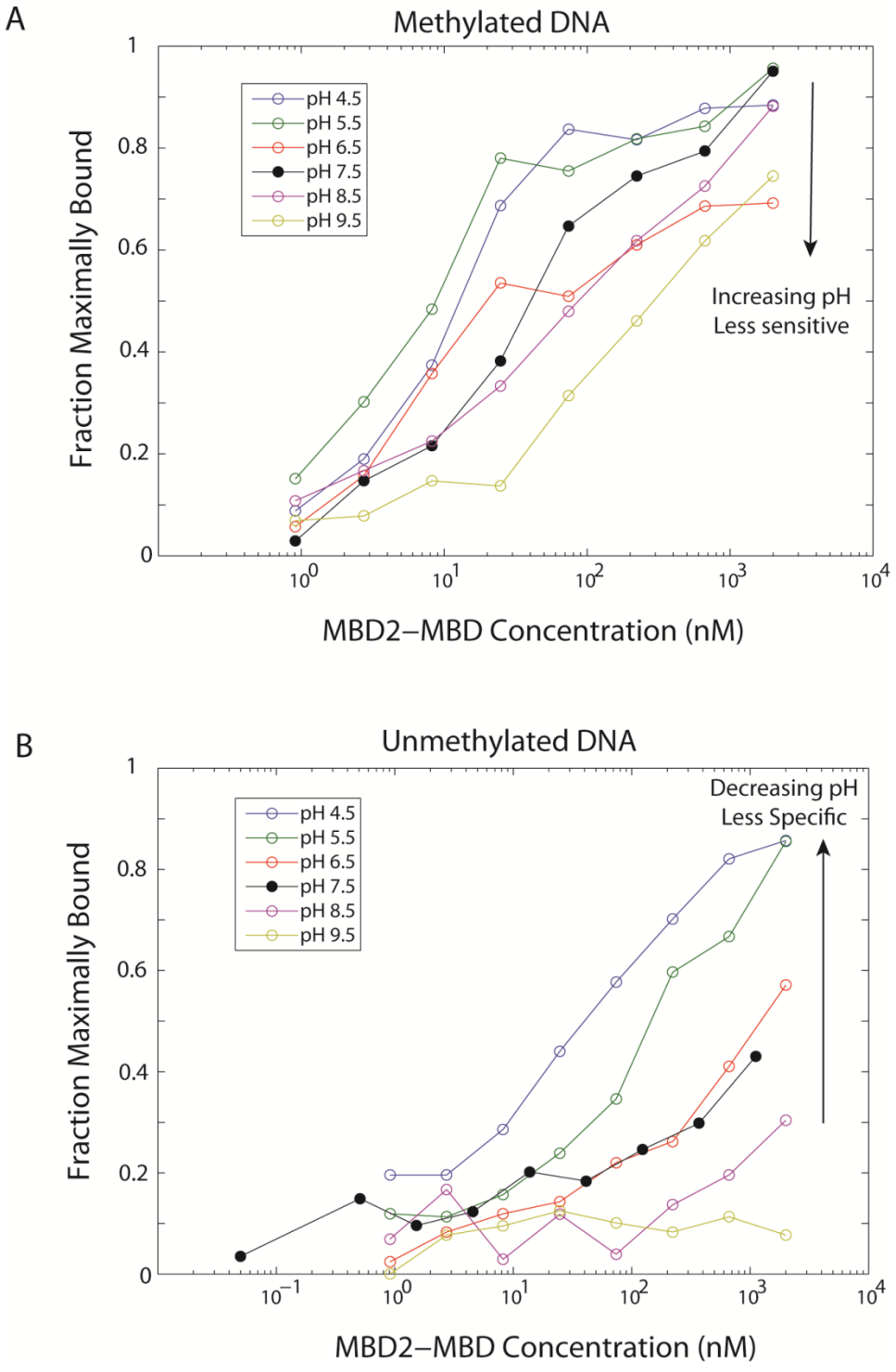


Figure 2.4: MBD2-MBD binding to DNA at different pH's

(A) MBD2-MBD binding to methylated DNA as measured by fluorescence polarization. Lower pH led to an increase in sensitivity **(B)** MBD2-MBD binding to unmethylated DNA. Higher pH led to higher specificity (less unmethylated DNA binding).

The final parameter queried was glycerol concentration. The intracellular environment is extremely rich in proteins and lipids. No association between MBD2 and lipids has been demonstrated, however, we wanted to confirm that its activity is not altered by the presence of detergent like materials. Addition of 4% or less glycerol did not appear to affect the ability of MBD2-MBD to bind methylated DNA (Figure 2.5). Concentrations above this level exhibited a moderate loss of sensitivity.

Isothermal Calorimetry Experiments to Confirm Binding Conditions

To confirm our results and finalize the optimal conditions for binding we performed isothermal calorimetry binding curves on the MBD2-MBD and meDNA interaction. Utilizing the same oligo and protein from our fluorescence polarization assay we tested low (50mM), optimal (125mM), and high (250mM) salt conditions (Figure 2.6A-B, Figure 2.7). Lower salt, as previously seen, showed nonspecific binding at higher protein concentrations in both hemi-methylated and unmethylated DNA substrates (Figure 2.8A). The nature of this interaction can be inferred by looking at the thermodynamics of the low salt meDNA-protein curve. The first binding event is entropically driven ($\Delta S=137$), consistent with the MeCP2 crystal structure which describes organized water as a key component of the MBD protein-methylated DNA interface. The second event, which overlaps with unmethylated and hemi methylated DNA

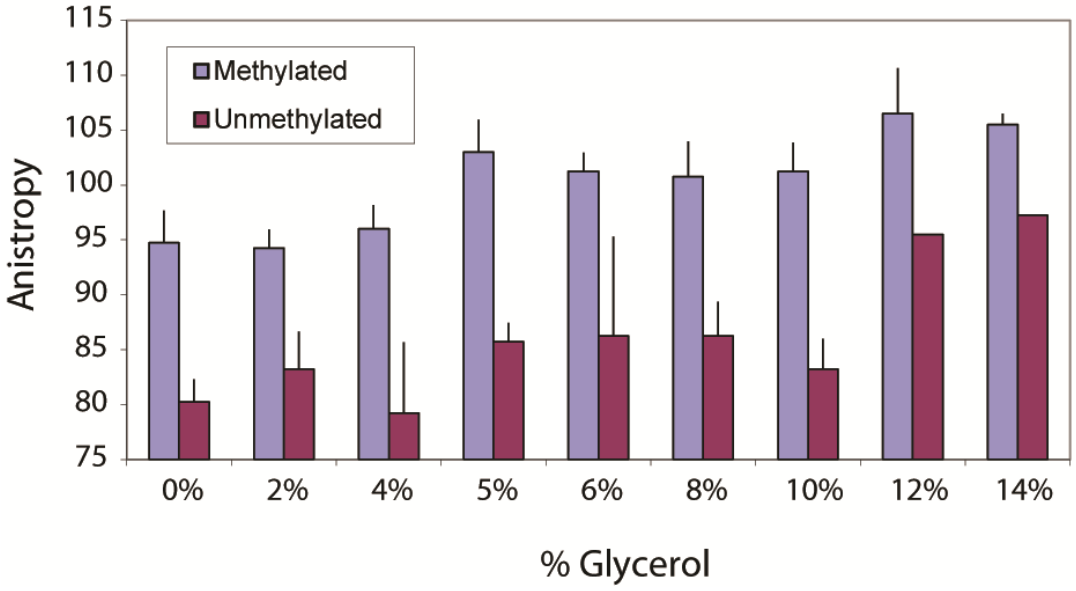


Figure 2.5: MBD2-MBD binding to DNA at different glycerol concentrations

Unmethylated and methylated DNA binding measured by fluorescence polarization assay in triplicate.

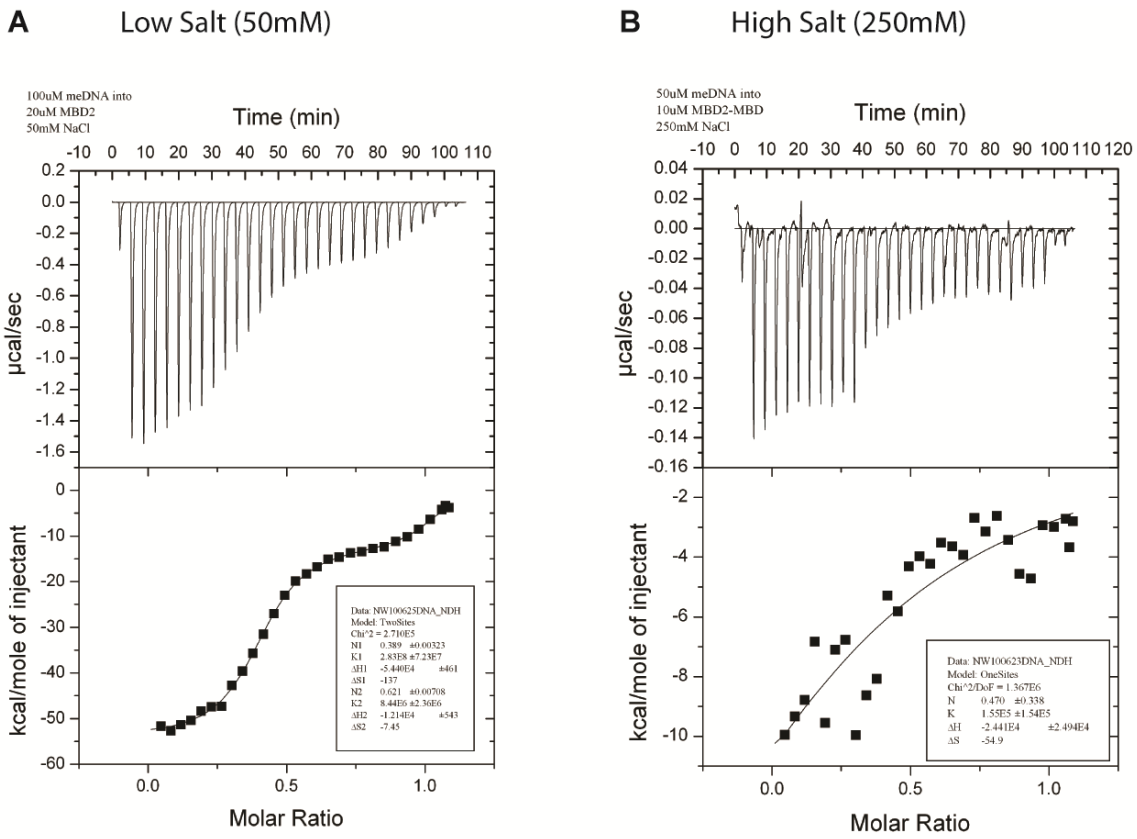


Figure 2.6: Isothermal calorimetry binding curves of MBD2 and methylated DNA at low and high salt

(A) ITC binding curve of MBD2-MBD with methylated DNA at low salt (50mM). **(B)** and high salt (250mM).

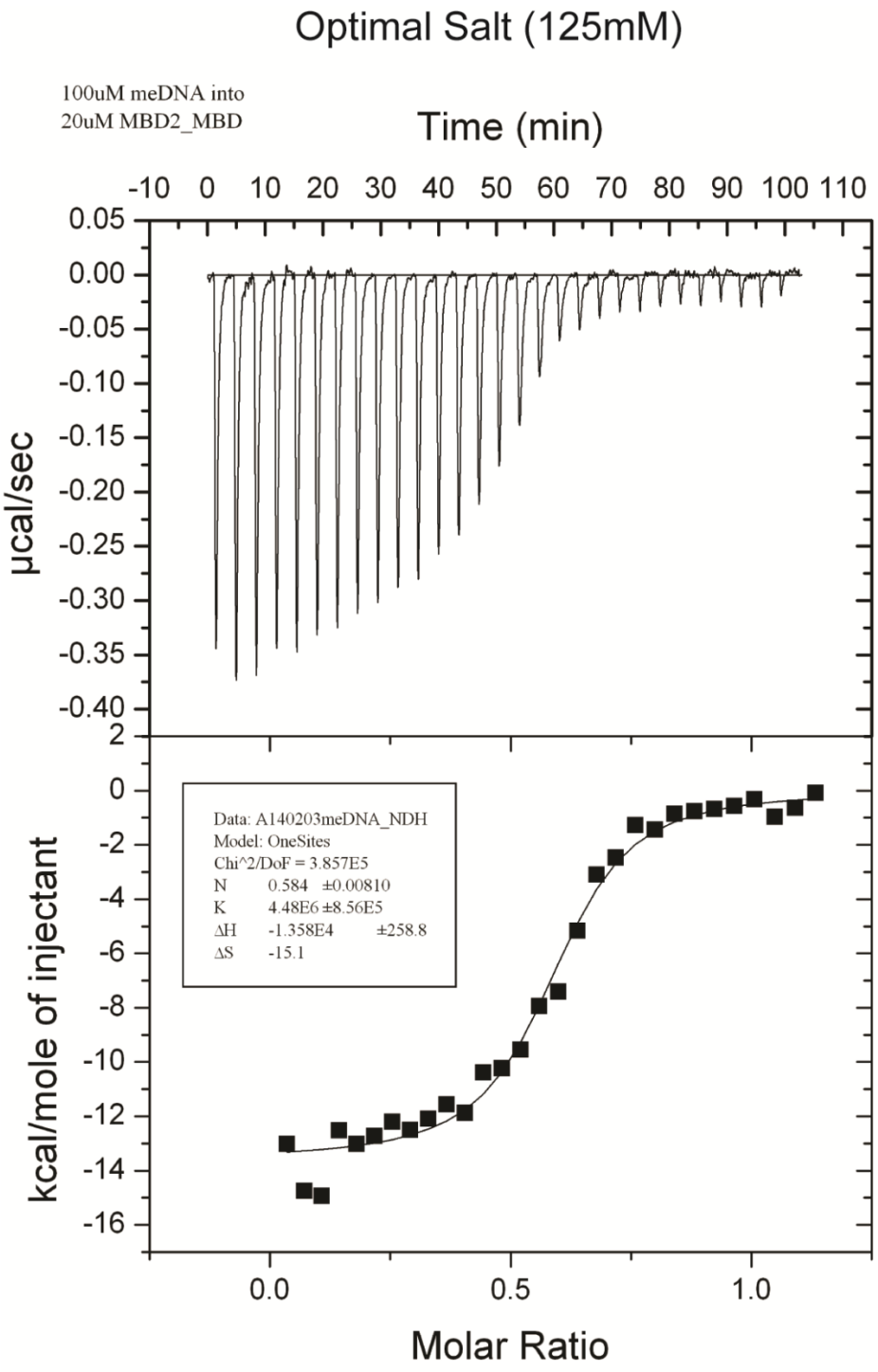


Figure 2.7: Isothermal calorimetry binding curve of MBD2-MBD with methylated DNA at optimal salt.

At optimal (125mM) salt and buffer conditions.

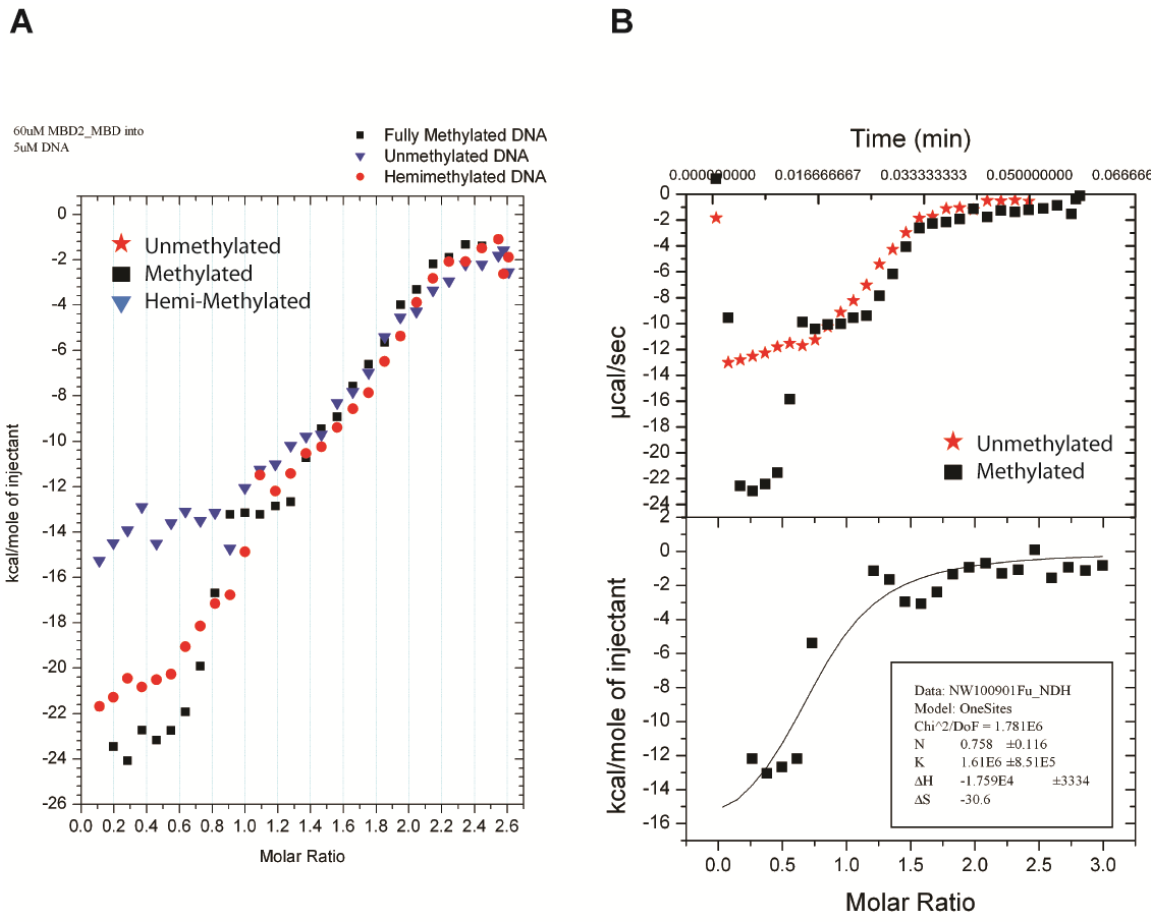


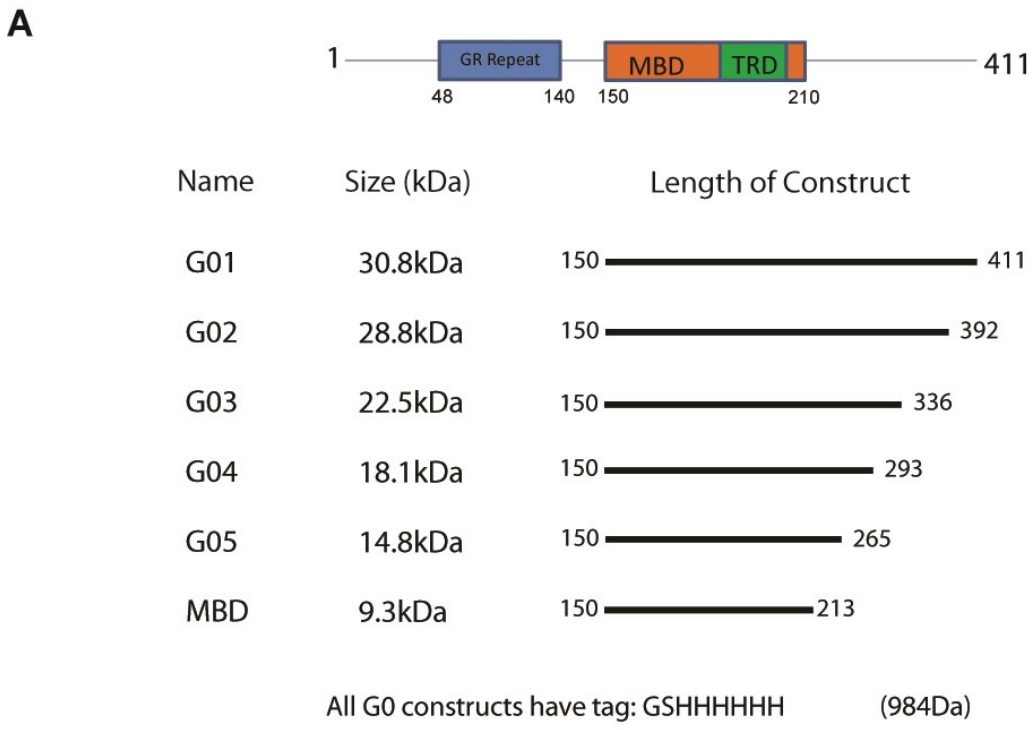
Figure 2.8: Fitted ITC curve of MBD2-MBD protein and DNA

(A) Methylated, hemi-methylated and unmethylated DNA binding curves with MBD2-MBD at low salt (50mM). **(B)** Methylated binding curve with unmethylated (background) heat subtracted.

binding curves, is more enthalpy driven, consistent with a protein seeking to mitigate an unmatched charge. High salt lowered sensitivity significantly (Figure 2.6B). The K_a for optimal conditions was $4.48x10^6$ M^{-1}, similar to the value obtained from the low salt when unmethylated DNA binding is subtracted away from the fully methylated values, $1.61x10^6$ M^{-1} (Figure 2.8B). Based on these results we identified the optimal pH (7-7.5), glycerol (4%), and salt concentration (125mM) for future work with the MBD2-MBD protein.

Biophysical Behavior of Longer MBD2 Constructs

To this point only the methyl binding domain of MBD2 had been utilized for experiments. To determine whether other portions of the protein impact binding we expressed 5 larger constructs in bacterial cells, containing progressively more of the C-terminus (Figure 2.9A). The first 150 residues of MBD2 encode the GR repeat domain. This region is predicted to be a coiled-coiled, and its presence has made *in vitro* purification difficult in the past (Reichert and Lee et al, unpublished). As such we opted to remove it from all constructs. After successful purification and storage we tested the various proteins' ability to bind fully methylated, hemimethylated and unmethylated 5-methylcytosine DNA using our fluorescence polarization assay (Figure 2.9B and 2.10). Unfortunately



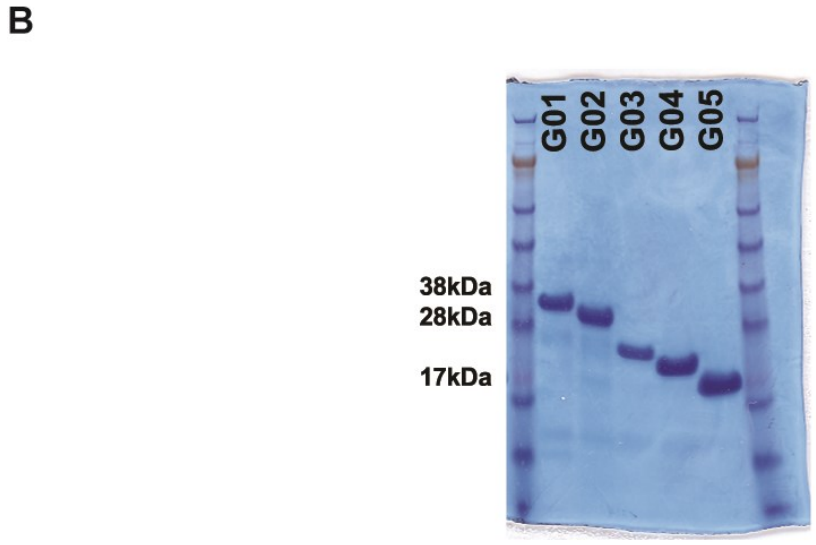


Figure 2.9: Construction, expression and purification of larger MBD2 constructs

(A) Layout of longer MBD2 constructs. **(B)** Coomassie gel on purified recombinant MBD2 constructs.

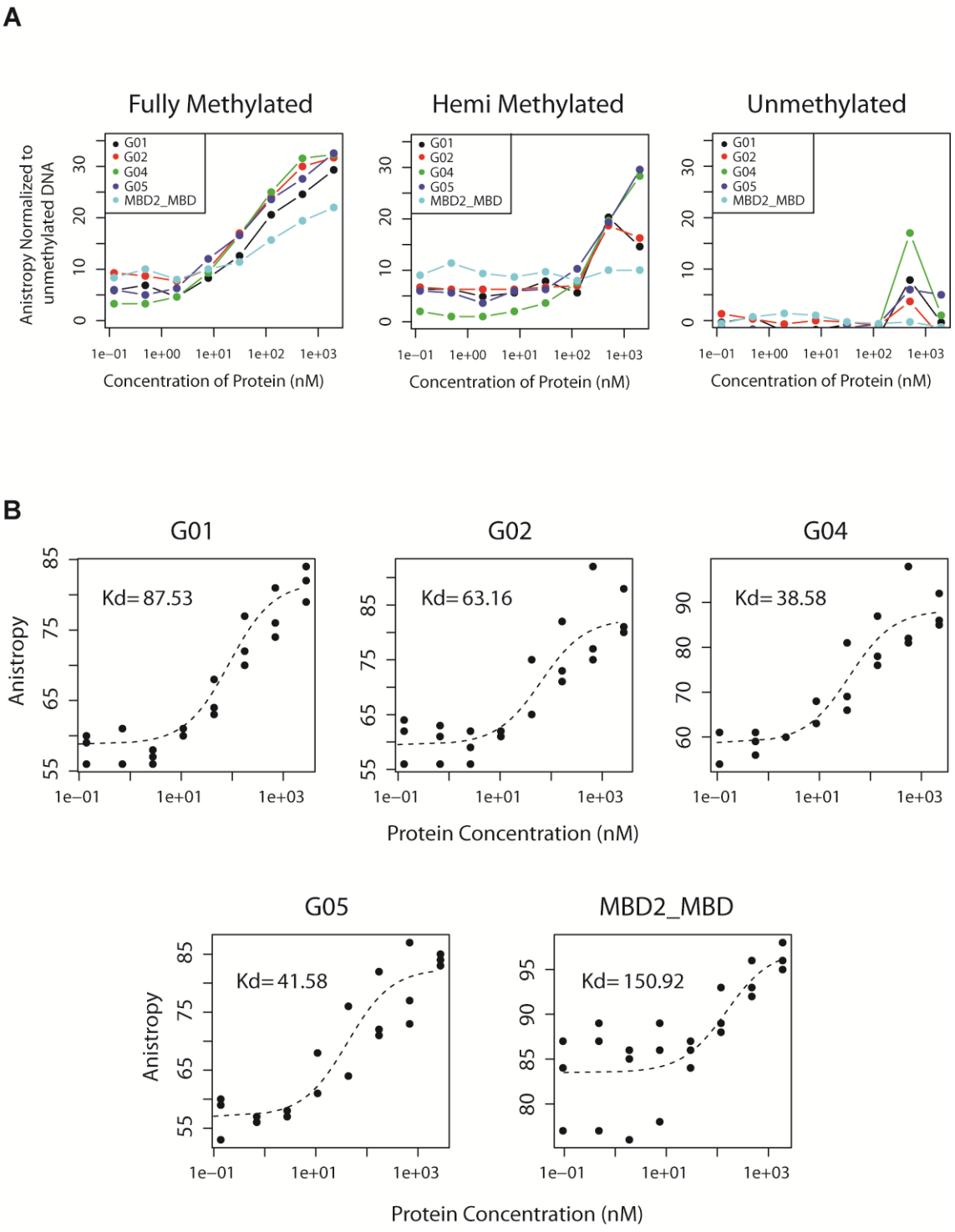


Figure 2.10: Binding curves for larger MBD2 constructs with DNA

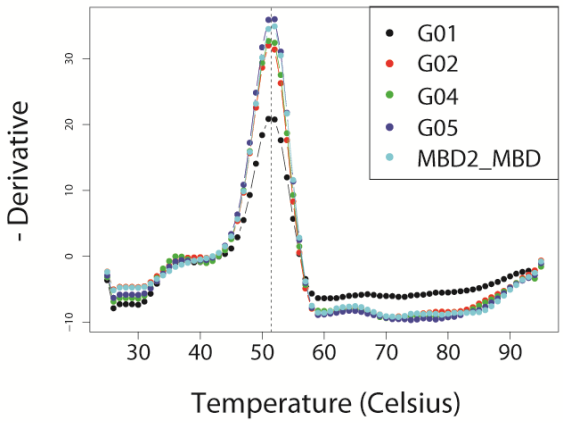
(A) Binding curves of longer MBD2 constructs with hairpin oligos containing a single symmetric CpG at with methylation states. Binding was measured using a fluorescence polarization assay. **(B)** Fitted binding curves for MBD2 constructs with fully methylated DNA. Approximate K_d's are shown for each protein.

MBD2_G03 did not express well and thus was not included in our initial analysis of these proteins. It was successfully expressed and purified later from bacteria and had almost identical binding results to those presented (results not shown). The larger constructs showed similar disassociation constants (K_d≈50nM) compared to the MBD domain only, but some of the larger constructs did appear to exhibit higher affinity for methylated DNA (Figure 2.10). Fully methylated DNA was the preferred substrate for all constructs with almost no binding present in the setting of unmethylated DNA, and only slight binding to asymmetrically hemi-methylated DNA (Figure 2.10A). These results suggest that the MBD function of MBD2 is modular, in that the domain is able to recognize and bind to methylated DNA in a specific manner with only some minor contribution to affinity from other domains or parts of the protein.

To characterize these proteins further we performed thermal melt curve and limited proteolysis experiments. Protein melt curves on each construct were conducted in the normal storage buffer (150mM NaCl, 10mM Tris pH 7.4, 1mM TCEP) and showed no difference between the proteins, with denaturing occurring at approximately 51.5^{o}C (Figure 2.11A). To confirm these results and determine if a more optimal buffer or pH existed we repeated the thermal melt experiment under varying conditions using the MBD2-G01 construct (Figure 2.11B-C). Lower pH (between 6.0-7.0) appeared to modestly stabilize the protein, with a 2-3^{o}C

shift in melting T_m. We then assessed various salts using pH values of 5.0, 6.0 and 7.0 and found that the majority made no difference in

A



B

pH Adjustment Plate Setup

	Salt	1	2	3	4	5	6	7	8	9	10	11	12
A	0.2M NaCl	N/A	0.1 M Sodium Acetate pH 4	0.1 M Sodium Acetate pH 5	0.1 M MES pH 6	0.1 M MES pH 6.5	0.1 M HEPES pH 7	0.1 M HEPES pH 7.5	0.1 M HEPES pH 8	0.1 M Tris pH 7.5	0.1 M Tris pH 8	0.1 M Tris pH 8.5	0.1 M CHES pH 9
B													
C	0.2 M CaCl2												
D													
E	0.2 M MgCl2												
F													
G	0.01 M ZnCl												
H													

Salt Adjustment Plate Setup

	1	2	3	4	5	6	7	8	9	10	11	12
A	Formate	Formate	Formate	Formate	Formate	Formate	Formate	Formate	Formate	Formate	Formate	Formate
B	Sulfate	Sulfate	Sulfate	Sulfate	Sulfate	Sulfate	Sulfate	Sulfate	Sulfate	Sulfate	Sulfate	Sulfate
C	Citrate	Citrate	Citrate	Citrate	Citrate	Citrate	Citrate	Citrate	Citrate	Citrate	Citrate	Citrate
D	Water	Water	Water	Water	Water	Water	Water	Water	Water	Water	Water	Water
	200mM	400mM	600mM	800mM	200mM	400mM	600mM	800mM	200mM	400mM	600mM	800mM

C

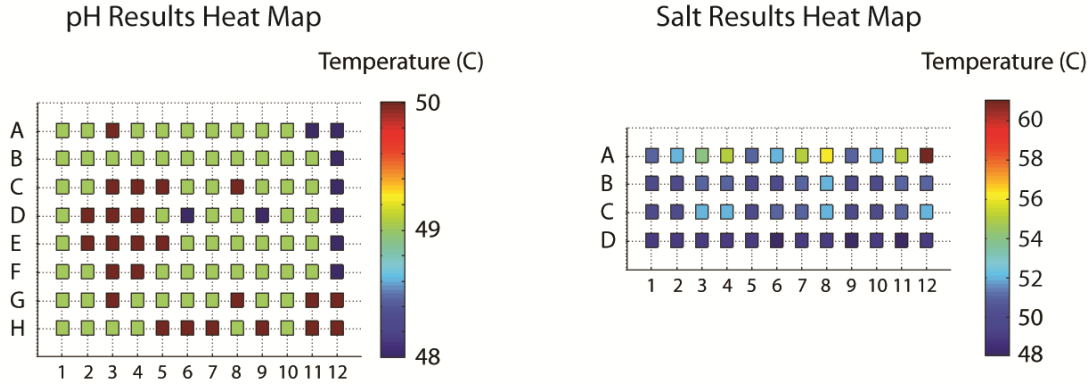


Figure 2.11: Protein thermal melt analysis on MBD2 constructs

(A) Protein thermal melt raw data for all MBD2 constructs. **(B)** Plate setups to test MBD2-G01 construct with thermal melt curves. **(C)** Results of MBD2-G01 thermal melt experiment under varying pH and salt conditions Protein is most stable at lower pH and with high formate.

protein stability. The exception was formate, which at concentrations of 400mM and above seemed to help stabilize the protein, as shown by a dramatic 8-10^{o}C shift in melting temperature (Figure 2.11C).

We next assessed our MBD2 constructs with a limited proteolysis experiment. Well folded proteins resist cleavage by proteases. Incubating the various constructs with a small amount of protease for a limited time should cleave only the peptide locations which are not well formed or stable, leaving only the domains/regions that are well packed. We exposed MBD constructs to subtilisin A, a serine protease with broad specificity, as well as trypsin, a commonly used serine protease that cuts after the basic residues lysine and arginine. After a 30 minute incubation the proteins were run on an SDS-PAGE gel and stained with Coomassie (Figure 2.12A) or silver stain (Figure 2.12B). Interestingly two relatively stable isoforms appeared in multiple proteins, indicated by arrows in the silver stain gel (Figure 2.12B). The larger is approximately the size of MBD2-G03, and the other approximately the size of MBD2-MBD. Excised bands sent for mass spectrometry analysis were inconclusive for the larger band, but confirmed the smaller one (lower arrow) as MBD2-MBD (results not shown).

Based on these results we looked more closely at the G01 (almost full length) and G03 constructs to assess the feasibility of using them for x-ray crystallography. First we attempted large scale affinity purification

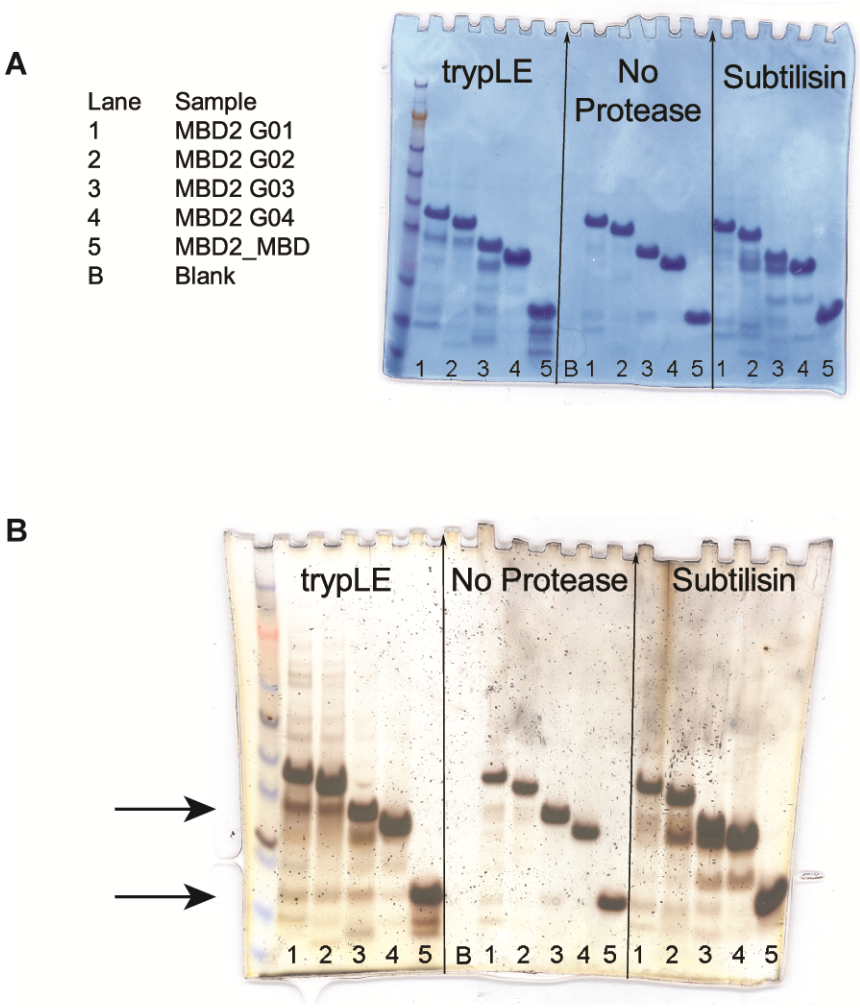


Figure 2.12: Limited proteolysis experiments with MBD2 constructs

(A) Coomassie and **(B)** silver stain gel of limited proteolysis with subtilisin A and trypsin on the larger MBD2 constructs. Arrows indicate two stable forms that appear in multiple constructs.

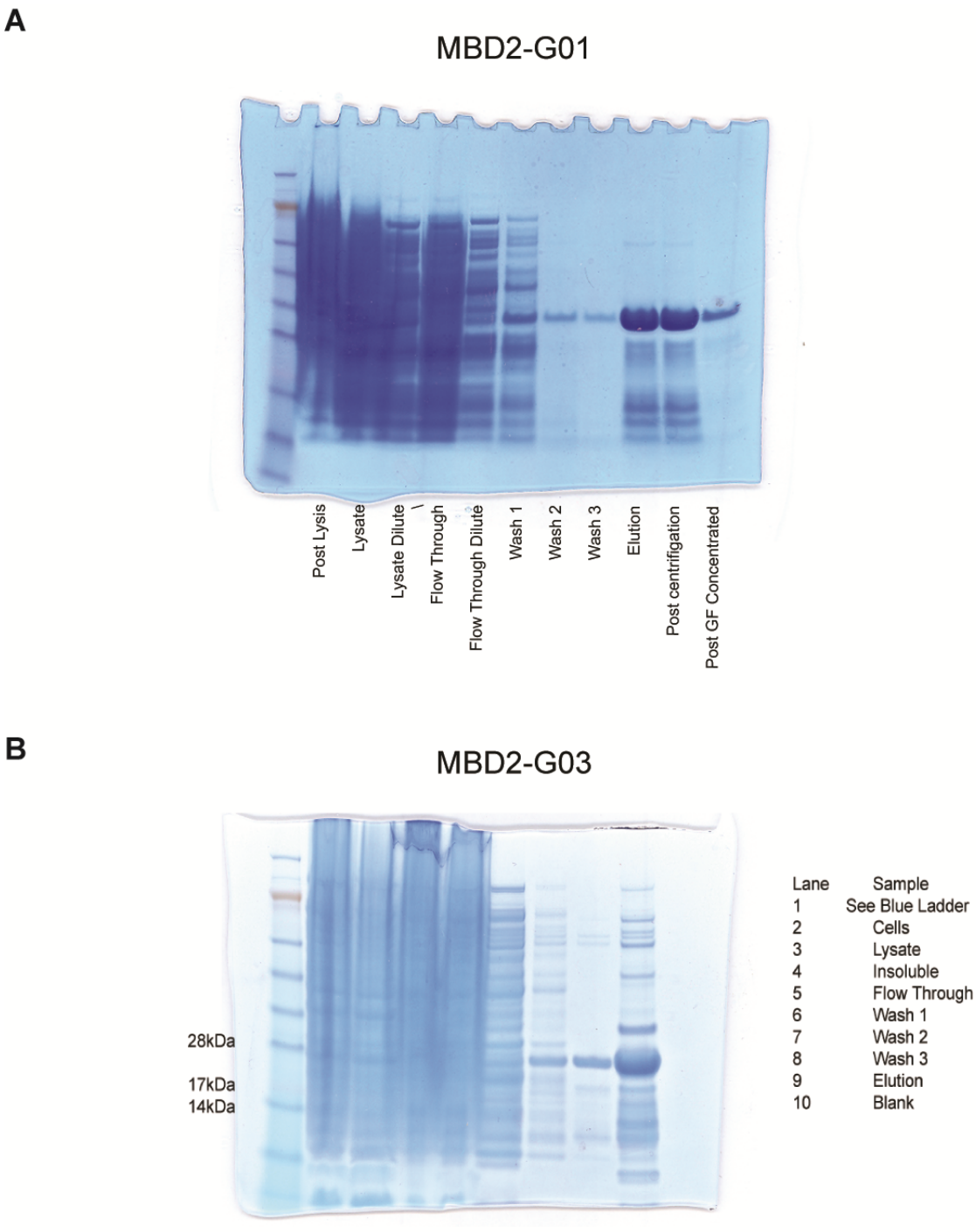


Figure 2.13: Large scale affinity purification of MBD2 constructs

(A) MBD2-G01 and **(B)** MBD2-G03 affinity purification with IMAC beads. The large band in each elution fraction is consistent with the recombinant protein of interest.

of the MBD2-G01 and G03 with the histidine tag tail on both proteins (Figure 2.13). We successfully purified both proteins but unfortunately MBD2-G01 became unstable during concentration and precipitated significantly. MBD2-G03, however, was amenable to concentrating and maintained purity >95% post size exclusion gel filtration purification (Figure 2.14A). Importantly, G03 was stable at lower concentrations of salt which would make screening crystal conditions easier (Figure 2.14B). Provided the above evidence we opted to move forward with MBD2-MBD and MBD2-G03 for structural efforts.

Crystallization of Native MBD2-MBD with Methylated DNA

MBD2-MBD was concentrated to ~20mg/mL (2mM) for initial structural work. The protein was screened by itself at 1mM using a 96 well hanging drop vapor diffusion approach with a total of five screens. After a week ~80% of wells showed precipitant with no crystals present. Despite repeat attempts at higher concentration and alternate ratios of mother liquor to protein we were unable to find a suitable condition with protein alone.

Provided the previous success of other groups with crystalizing MBD's with DNA we opted to add DNA to MBD2-MBD and rescreen. A total of six 96 well screens were performed using three different pre-

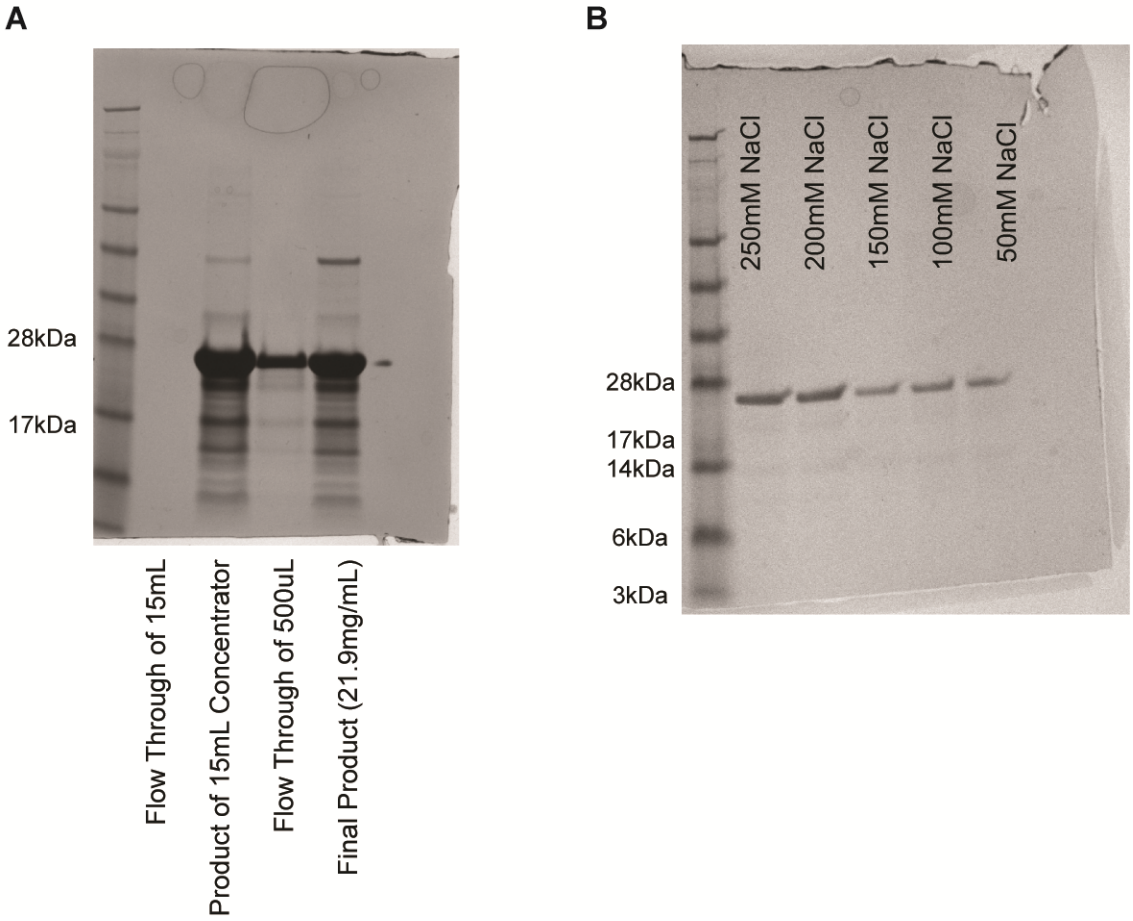


Figure 2.14: Gel filtration purification and stability studies with MBD2-G03

(A) Coomassie staining of MBD2-G03 gel filtration and concentration results. **(B)** Coomassie stain of MBD2-G03 dialyzed overnight into progressively lower salt concentrations. Loss of signal suggests protein is becoming unstable and precipitating.

annealed double stranded DNA sequences with overlapping A or T bases to allow concatemerization of adjacent strands. After approximately 5-6 days we observed large cuboid crystals in wells containing the MBD2-MBD with 14bp DNA sequence in a 2:1 ratio respectively. After optimization we achieved cuboid crystals of approximately 200μm x 200μm x 200μm size, but only in wells with ≥1.4M citrate and pH≤7 (Figure 2.15A). This is consistent with our thermal melt data which showed that MBD2-MBD is stabilized in lower pH and highly concentrated salts containing carboxyl functional groups (such as formate and citrate). Previous reports suggest that proteins crystallize more readily when in a stable form or condition(*103*). We cryoprotected crystals in 2.0M citrate and collected diffraction data on a local x-ray source (see methods for details). The crystal diffracted to approximately 2.2Å in a $P3_N21$ space group (Table 2.2). A single crystal was washed numerous times in citrate buffer and run on an SDS-PAGE gel, confirming it contained protein (Figure 2.15B).

With our collaborators we made numerous, but unsuccessful, attempts at solving the phase of the MBD2-MBD native crystals with the MeCP2-MBD crystal structure by molecular replacement (PDB accession number 3C2I). Heavy metal soaking was also attempted unsuccessfully, mostly due to human error, but also difficulties with the native crystals breaking apart when metals were introduced. Ultimately four

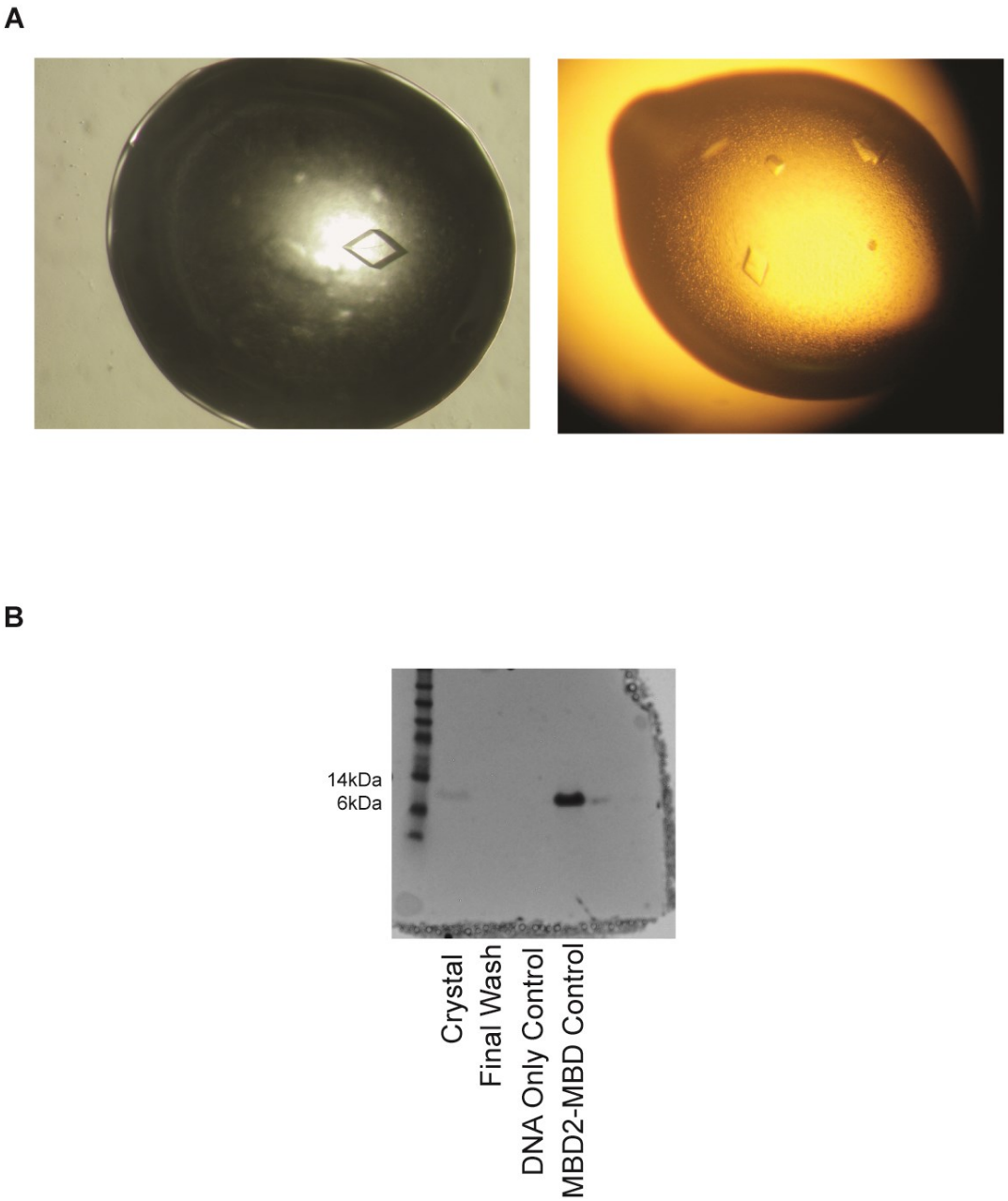


Figure 2.15: Native crystals of MBD2-MBD with 14bp

(A) Crystals formed by MBD2-MBD with 14bp DNA in a 2:1 ratio, respectively, after 3-6 days in 1.6M citrate pH 6.5. **(B)** Coomassie staining of a single crystal washed and run on an SDS-page gel.

Protein	**MBD2-MBD**	**MBD2-MBD-R195M**	**MBD2-MBD-L193M**
DNA	14bp meDNA	14bp meDNA	14bp meDNA
Screen Condition	1.4M Citrate pH 6.5	28% PEG8000 0.2M Sodium Acetate 0.1M Bis Tris Propane pH 6.5	2.0M ammonium sulfate 0.1M Na Acetate trihydrate, pH 4.6
Cryoprotectant	2.0M Citrate pH 6.5	42% PEG8000 0.2M Sodium Acetate 0.1M Bis Tris Propane pH 6.5	2.0M ammonium sulfate 0.1M Na Acetate trihydrate, pH 4.6 20% Glycerol
Space group	$P3_N21$	$P3_N21$	$P3_221$
a, b, c (Å)	52.592, 52.593, 110.524	56.027 56.027 110.063	52.304, 52.304, 110.141
α, β, γ	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	25-2.35 (2.43-2.35)	50-2.85 (2.95-2.85)	2.3
Rmerge (%)	5.8 (>100)	20.2 (1.2)	11.6 (>100)
I/σ	32.9 (1.43)	13.1 (1.0)	14.5 (1.2)
Completeness(%)	100.0 (100.0)	99.4 (99.6)	99.9 (100)
Redundancy	10.2 (8.9)	6.9 (6.1)	8.95 (6.65)
Rcryst (%)	N/A	N/A	In Progress
Rfree(%)	N/A	N/A	In Progress
Beamline	Rigaku FR-E x-ray with a Saturn 944+ CCD detector	APS 23-ID	APS Beamline 7-2
Dimensions (μm)	300x300x300	150x30x20	40x40x40
Phasing			
No. of Selenium Sites	N/A	N/A	3
Figure of Merit from SAD	N/A	N/A	0.29 (0.17)

Table 2.2: Summary of data collected on native and selenomethionine MBD2-MBD crystals

selenomethionine MBD2-MBD constructs were made as described in the methods. All of the endogenous methionines in MBD2-MBD are predicted to be in disordered regions, prompting us to mutate one into a predicted alpha helix. MBD2-MBD-G198M was unstable after purification and could not be concentrated to any appreciable level. MBD2-MBD-T200M did not express despite confirming by sequencing that the plasmid was in frame and contained the correct coding sequence. These two proteins were abandoned.

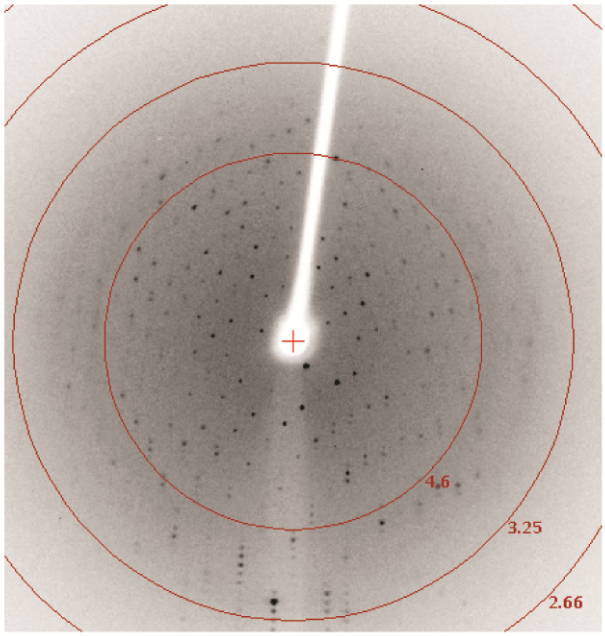
MBD2-MBD-R195M did express and we harvested crystals from two suitable conditions. The first was in 16-19% PEG3350, 0.2M Zinc Acetate and 1mM TCEP, cryoprotected in the same buffer supplemented with 20% glycerol. Single wavelength anomalous diffraction (SAD) data were collected at the advanced photon source at Argonne National Laboratory synchrotron but when phases were solved and initial density maps constructed it appeared only DNA was present. Review of our data suggests we accomplished phasing via a fixed zinc atom, which happens to contain an absorption edge very similar to selenium.

Repeat sparse matrix screening identified a second suitable condition: 30% PEG8000, 0.2M Sodium Acetate, 0.1M Bis Tris Propane pH 6.5. Multi wavelength anomalous diffraction (MAD) data were collected for selenomethionine at the advanced photon source at Argonne National Laboratory synchrotron (Table 2.2). Unfortunately, despite numerous attempts, we were unable to solve the phase with these data,

and suspect that the selenomethionine sites were not ordered enough to locate.

MBD2-MBD-L193M crystals were also successfully attained using the condition: 2.0M ammonium sulfate, 0.1M sodium acetate trihydrate pH 4.6. After cryopreservation SAD data were collected and processed (Figure 2.16A, Table 2.2). With our collaborators we successfully located 3 selenomethionine sites with this data set and overlay of the MeCP2-MBD solution on initial electron density maps suggests there is density consistent with both DNA and protein present (Figure 2.16B). Work is still underway to obtain a final solution for this structure.

A



B

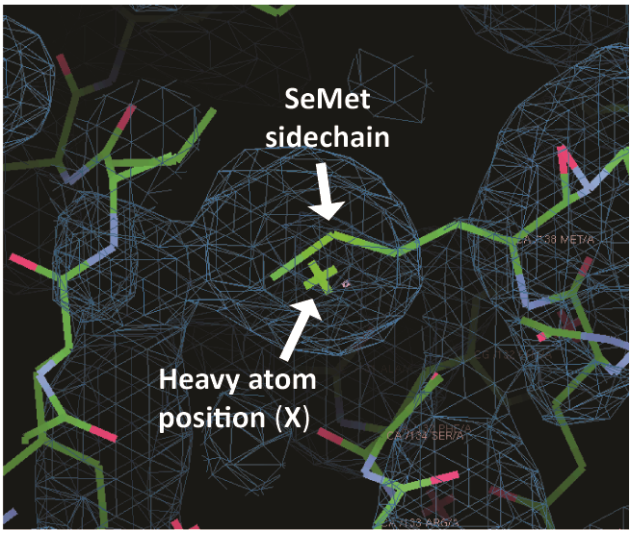


Figure 2.16: Data collection and analysis of MBD2-MBD-L193M selenomethionine crystals

(A) Diffraction of MBD2-MBD-L193M selenomethionine crystals during screening at JHMI in preparation for a synchrotron trip. **(B)** Representative example of 1/3 selenomethionine sites located in the initial electron density maps obtained from SAD data. The images provided in this figure were provided by Pat Byrnes in the lab of Dr. Dan Leahy.

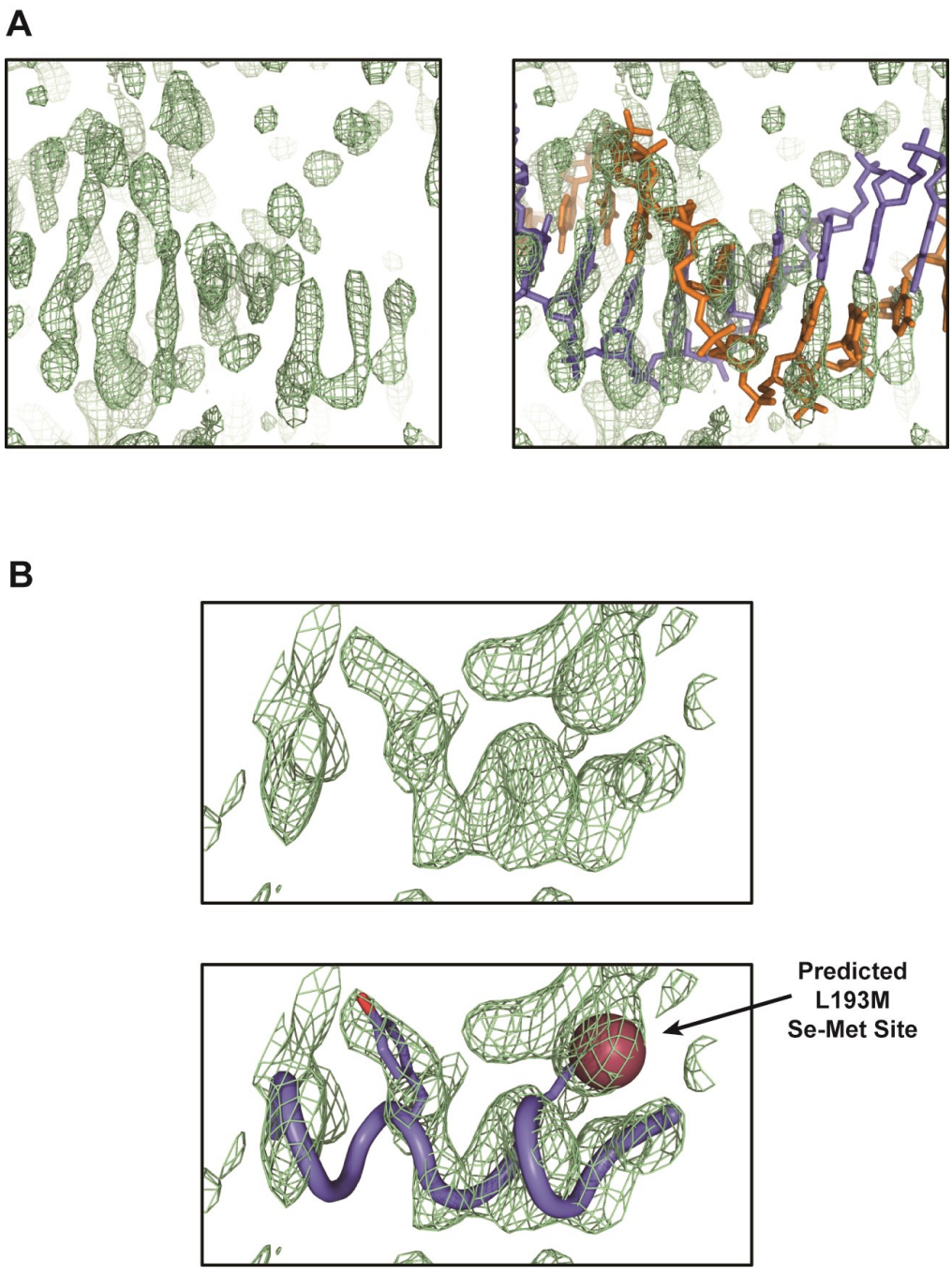


Figure 2.17: Initial electron density maps and model with MBD2-MBD-L193M SAD data

(A) Electron density consistent with DNA backbone. The left and right figures are the same, with generic DNA modeled on the right. (B) Electron density consistent with protein. In the bottom panel the MeCP2 model is overlaid and the predicted site for the se-met labeled L193M does overlap with an identified selenium site. Images were kindly provided by Patrick Byrnes.

2.4 Discussion

Despite over a decade of investigation we still do not completely understand the mechanism behind the MBD family member interaction with methylated DNA. A lot of effort investigating MeCP2 has occurred, presumably because it was the first member identified and the only one with a relevant human disorder associated with its mutation. In the work described here we attempt to fill the gap and answer the relevant biophysical questions regarding MBD2's MBD.

How MBD's are targeted to the genome is an extremely relevant and interesting problem. Our data suggests that by changing salt and pH conditions the specificity and sensitivity can be drastically changed. Accomplishing this inside a nucleus *in vivo*, where MBD2 is found, is quite difficult and thus an unlikely mechanism. This data has assisted us, however, with establishing optimal conditions for using the MBD2-MBD as a tool to enrich methylated DNA fractions in an *in vitro* setting. By increasing salt and using a slightly higher pH we've been able to improve specificity with only a moderate loss in sensitivity.

The differences between insect and bacterial protein binding could, however, suggest a mechanism by which the cell can alter MBD2 function. The post translational modifications present from the SF9 cells do seem to improve binding by gel shift. While this experiment was not investigated further it could, along with other evidence in the literature, provide insight into a mechanism for controlling MBD binding(*102*).

MBD2 targeting kinases could localize to particular places in the genome and thereby lower the chances of MBD2 effectively binding to adjacent genomic locations. Clearly more work should be done in this area to establish the relevant post translational marks and their function.

The MBD2-MBD crystallography work was a large part of my studies of this protein. Despite numerous initial failings with the protein alone, complexes with methylated DNA in both our hands and others seem to produce diffraction worthy crystals under many different conditions. It is troublesome that no one to date has successfully crystallized an MBD family member in its entirety or with just the MBD protein alone. While there has yet to be a conclusive experiment exploring this question we speculate it may be due to the protein's inherent stability/folding in solution. We suspect that, when unbound, MBD2, in particular the MBD, remains relatively loosely folded in its tertiary structure. Our gel filtration studies support this theory, as the protein rarely elutes in one uniform peak. In the presence of methylated DNA, however, MBD2 achieves a more rigid structure as favorable contacts emerge between the protein and DNA. This more stable form likely initiates enzymatic activity through the MeCP1 complex. Again gel filtration and other experiments lend some evidence to this idea, with premixed protein-meDNA constructs running at one uniform band or peak during purification. The limited proteolysis experiments also support this theory, with numerous smaller peptides appearing in a

short incubation time in our gels. Crystals, by their nature, require a rigid symmetry to form, and thus a loosely folded protein presents a difficult substrate to crystalize. More investigation into this area could help decipher the mechanism behind MBD signaling to other complex members and/or conditions under which an MBD may crystalize on its own. Such information could be helpful for structure aided design if suitable inhibitors are found.

Despite the significant effort described here there remain a number of unanswered questions regarding the mechanisms of MBD family member interaction with methylated DNA. We look forward to solving the MBD2-MBD crystal structure in the near future and continuing investigations into biophysics of the MBD2 protein.

2.5 Methods

Comparison of Recombinant MBD2-MBD from Bacterial and Insect Cells

A codon optimized sequence for the MBD2-MBD polypeptide was synthesized and cloned into the pGSE6 vector (Genscript USA Inc) for expression in bacteria as a C-terminal hexa-histidine tagged fusion protein. Briefly, BL21-DE3 cells (Agilent Technologies) were transformed with this construct, allowed to grow to an OD_{600} of 1.0, and were induced with 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG, Corning Cellgro) overnight in a shaking incubator at 220 rpm and 20°C. Bacteria were lysed using a French press in equilibrium buffer containing 300 mM NaCl, 50 mM sodium phosphate pH 8.0, 5 mM imidazole (Sigma Aldrich), and a protease inhibitor cocktail (Roche USA). Lysates were centrifuged at 17,000x*g* for 30 minutes and then incubated with IMAC Nickel NTA beads (Bio-Rad) for 1 hour at 4°C. The beads were washed three times with equilibrium buffer containing 15 mM, 20 mM and 25 mM imidazole, respectively, and affinity purified protein was eluted in equilibrium buffer containing 150mM imidazole. The his-tagged protein was further purified by gel filtration using a Superdex g75 26/300 Column (GE Healthcare Life Sciences) in storage buffer (150mM NaCl, 20mM Tris pH 7.4, 1mM tris(2-carboxyethyl)phosphine (TCEP, P212121 LLC, Toledo, Ohio). Fractions containing the purified MBD2-MBD were concentrated with 3000 MWCO Amicon® Ultra-4 Centrifugal Filter Units (Millipore) and flash frozen with liquid nitrogen and stored at -80°C.

MBD2-MBD from insect cells was made using the same primary sequence as the MBD2-MBD in bacterial cells. Briefly, 2-4mL of previously designed and produced baculovirus (T. Speed and Z. Reichert) was used to infect 2.2e6 SF9 cells/mL (ATCC) in a 250mL plastic flask (Corning) containing SF900III media (Invitrogen). Cells were incubated for 48 hours at 28^{o}C and 130rpm and harvested by centrifuging at 1000xg for 10 minutes, then snap frozen in liquid nitrogen. Protein purification followed the same protocol listed for the bacterial cells with the following exception: Lysis was performed by 3 repeated freeze (liquid nitrogen) thaws (42^{o}C water bath). The lysate was then run through a 20ga needle via a syringe 3 times. Clearing of the lysate and protein extraction continued as described above.

Gel Shift Assay

For the gel shift assay DNA hairpin oligo with sequence ATGCTC^{me}GTAGCACTTTTGTGCTAC^{me}GAGCAT (IDT) was annealed by mixing heating to 95^{o}C for 5 minutes, followed by rapid cooling by moving directly to ice until reaching room temperature. Once annealed oligos were labeled with P^{32} (Perkin Elmer) by mixing 2.5µL of 10µM oligo with 2.5µL of x10 T4 PNK buffer (Promega), 2.5µL of T4 Polynucleotide Kinase enzyme (Promega), 9µL P^{32} at 250µCi, and water to a total volume of 20µL. The mixture was incubated for 1 hour at 37^{o}C, then at 65^{o}C for 20 minutes. Unbound radioactive phosphate was removed by running

the mixture through a G-25 gel filtration spin column (GE). The eluted product was diluted 1:25 to a concentration of 40nM and stored at -20°C. For the assay, 5µL of oligo was incubated with increasing concentrations of MBD2-MBD protein (from 0.1µM to 5µM) in the presence of x1 gel shift buffer (Promega) and water up to a total volume of 20µL No protein served as the negative control. This was incubated for 30-60min at room temperature and run on a 20% TBE agarose gel (Invitrogen). Gels were vacuum dried for 1hour (room temperature or 70°C) and exposed to developing film. For the NEM (Sigma Aldrich) treated samples the same procedure was followed, except 5mM NEM was added during the protein and meDNA incubation step.

MBD2-MBD Binding Under Different Physiological conditions

Fluorescence Polarization Binding Assays

Fluorescence polarization binding assays were performed as previously described(*63*). Briefly, MBD2-MBD protein was added in a dilution series to the reaction buffer containing varying levels of glycerol (0-14% in 2% steps), NaCl (50-200mM in 25mM steps) or buffer (adjusted to appropriate pH 4.5-9.5 in 1.0pH steps) along with 1 mM $MgCl_2$, 0.5 mM EDTA, 0.5 mM DTT, 0.2% Tween-20 up to a total working volume of 50µL. For the pH adjusted experiments MES (Sigma Aldrich) was used for pH adjustments from 4.5-6.5, and Tris (Sigma Aldrich) was used for pH 7.5-9.5. Hairpin-forming oligonucleotides with the sequence 5′-

fluorescein (FAM)-ATGCTCGTAGCACTTTTGTGCTACGAGCAT-3′ (unmethylated) or 5′- fluorescein - ATGCTC^{me}GTAGCACTTTTGTGCTAC^{me}GAGCAT-3’ (methylated) were annealed by heating to 95^{o}C and cooling on ice rapidly. They were then added to the binding reaction to a final concentration of 10nM. Binding assays were carried out in 96-well black half area plates (Corning) in triplicates. The plate was incubated at 4^{o}C for 1 hour with gentle shaking. Fluorescence polarization readings were performed on a Safire^{2} (Tecan) instrument with excitation at 470nm and monitoring emission at 525nm. Anisotropy values were acquired directly from the machine readout and plotted relative to MBD2-MBD protein concentration and fitted using the open source software R version 2.12. Z’ factors were calculated using the equation (1):

$$Z' = 1 - 3\,x\,\frac{SD_m - SD_u}{|\mu_m - \mu_u|} \qquad (1)$$

where SD_m and μ_m are the standard deviation and mean anisotropy values respectively of protein binding to the methylated FAM labeled oligonucleotide, and SD_u and μ_u are the standard deviation and mean of anisotropy values respectively of the protein binding to the unmethylated FAM labeled oligonucleotide.

Isothermal Calorimetry Binding Experiments

For isothermal calorimetry experiments recombinant bacterial made MBD2-MBD was placed into an ITC buffer containing 250mM,

125mM or 50mM NaCl (Sigma Aldrich), 1mM $MgCl_2$ (Sigma Aldrich), 0.5mM EDTA (Sigma Aldrich), 0.5mM TCEP (P212121) and 10mM Tris pH 7.4 (Sigma Aldrich). Protein was diluted to either 10 or 20μM in a final volume of 2mL. For the injector, methylated hairpin DNA from the fluorescence polarization experiment described above was used. The meDNA was mixed with the same ITC buffer described above to a final concentration of 50 or 100μM with a final working volume of 500μL. Both the DNA and protein mixtures were degassed for 5 minutes before carefully being placed into either the syringe (meDNA) or bomb (protein) ensuring no bubbles were present, all per the manufacturer's instructions. A MicroCal VP-ITC machine (GE) was used for all experiments. Data was collected using the following settings: inject volume 9μL, 30 total injections, time between injections 210 seconds, temperature 25^{o}C, stirring speed 270rpm, and a 10μCal/sec baseline heat. Data was analyzed and fitted using the Origin® software provided by MicroCal, utilizing a chi squared statistic to determine optimal fitting parameters for a 1 site or 2 site binding event.

For comparing MBD2-MBD affinity to different DNA substrates the process was the same, but only 50mM NaCl was used. For these experiments the protein as diluted to 60μM in 500μL working volume and placed in the syringe, while the appropriate DNA (either fully, hemi or unmethylated) was diluted to 5μM in 2mL working volume and placed in the bomb. Analysis was the same as described above.

Biophysical Behavior of Longer MBD2 Constructs

Purification of MBD2 Longer Constructs

Larger MBD2 constructs were purified following the similar methods listed above under the comparison of bacterial and insect protein section for MBD2-MBD from bacteria. Expression constructs were PCR amplified following the recommended protocol in the Fail Safe product manual (Epicentre) from a bacterially codon optimized vector made by GenScript (New Jersey) (Table 2.3). The PCR products were topo cloned following recommended protocols into a p-EXP-5CT vector (Invitrogen) containing a C-terminal hexa-histidine tag. These plasmids were transfected into BL21-DE3 cells (Agilent Technologies) and the bacteria allowed to grow to an OD_{600} of 1.0 in Terrific Broth (Corning-Cellgro). A total of 20mL of bacteria for each construct was induced with 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG, Corning Cellgro) and grown overnight in a shaking incubator at 220 rpm and 20^{o}C. The protein was purified using the Talon 20mL Single Step Purification kit (Clontech) using their recommended buffers and incubations. Protein

Name	Sequence
Forward	AGTGAAttcCATATGCGTGC
G01 Reverse	AAGCTTTCAGTGGTGGTGGTGGTGGTGAGAACCggcttcatcgcccgaatc
G02 Reverse	AAGCTTTCAGTGGTGGTGGTGGTGGTGAGAACCcaggatatccgccatcag
G03 Reverse	AAGCTTTCAGTGGTGGTGGTGGTGGTGAGAACCaatcggtgcggaactggt
G04 Reverse	AAGCTTTCAGTGGTGGTGGTGGTGGTGAGAACCggaggcactcagaccctg
G05 Reverse	AAGCTTTCAGTGGTGGTGGTGGTGGTGAGAACCattgctcggatggttcgt

Table 2.3: Primers used to create MBD2 longer constructs.

Amplified from a bacteria codon optimized construct containing the full length MBD2 protein.

was then concentrated in a 15mL volume 10,000 MWCO Amicon® Ultra-4 Centrifugal Filter Units (Millipore) and stored in storage buffer (150mM NaCl, 10mM Tris pH 7.4, 1mM TCEP). Protein was flash frozen in liquid nitrogen and stored at -80^{o}C.

Fluorescence Polarization Binding Curves of MBD2 Constructs

Fluorescence polarization binding curves were performed as previously described in the comparison of bacteria and insect MBD2-MBD section. The larger constructs were substituted in place of MBD2-MBD in the appropriate wells. The assay buffer contained 4% glycerol, 1mM $MgCl_2$, 0.5mM EDTA, 0.5mM DTT, 125mM NaCl, 10mM Tris HCl (pH 7.4), 0.2% Tween-20 (all from Sigma Aldrich) with a final working volume of 20µL. Oligos were hairpins with identical base pair sequences, methylation status being their only difference:

Methylated ATGCT^{me}CGTAGCACTTTTGTGCTA^{me}CGAGCAT

Hemi-methylated ATGCT^{me}CGTAGCACTTTTGTGCTACGAGCAT

Unmethylated ATGCTCGTAGCACTTTTGTGCTACGAGCAT

Primers were annealed prior to use by heating to 95^{o}C and then rapidly cooled on ice to room temperature. Oligos were used a final concentration of 10nM in the assay. Protein was assayed from 2000nM in a 4 fold decreasing fashion to about 0.12nM. Black 96 well half area plates (Corning) were used for all assays. Readings were performed using a Safire^{2} (Tecan) instrument with excitation at 470nm and monitoring

emission at 525nm. Anisotropy values were acquired directly from the machine readout and plotted relative to MBD2-MBD protein concentration and fitted using the open source software R version 2.12.

Thermal Melt Analysis of MBD2 Constructs

Thermal melt curves were performed in 96 well PCR plates (Bio-Rad) in a total working volume of 20µL. Approximately 2.5µg of protein was added to each well in a buffer containing 10x Sybr Orange (Life Technologies), and 50mM NaCl and 10mM Tris pH 7.4. Proteins were heated on a CFX96 real time PCR machine (Bio-Rad) from 25^{o}C to 95^{o}C in 0.5^{o}C steps with a 5 second hold at each step. Analysis and heatmaps were created using Matlab® and software/protocols described online in an open source website developed by Johns Hopkins School of Medicine researchers ((*103*), http://www.hotproteins.com/).

For assessing protein stability in pH and salt buffers the same final concentrations of material described above was used, but was added in 2µL to each well (added at 10x concentration). The remaining 18µL of the working volume consisted of the buffer or salt being tested at the described concentration (Figure 2.11B). Heating and reading was performed as described above.

Limited Proteolysis Experiments

Previous work (not described) tested protease concentrations of 10ng, 1ng, and 0.1ng to determine the optimal concentration for the MBD2 protein. The optimized protocol used 10µg of protein construct added to a 12 well strip tube (USA Scientific). This was diluted with 100mM NaCl, 20mM Tris pH 8.0 and 1mM β-Mercaptoethanol (Sigma Aldrich) to a final volume of 16µL. 0.1ng of TrypLE (Cellgro) or subtilisin A (Sigma Aldrich) was added to each well and mixed with pipetting at a final working volume of 20µL. Samples were incubated at 37°C for 30 minutes. 6µL x4 LDS loading buffer with SDS (Invitrogen) was quickly added to each well and then heated to 95°C to stop protease activity. Samples were run on a SDS-PAGE gel (Invitrogen) and visualized using SimplyBlue™ (Invitrogen) or silver staining (Bio-Rad) following manufacturer instructions. Bands were excised using a straight razor and stored in 1.5mL Eppendorf tubes. They were then submitted to the Johns Hopkins Mass Spectrometry Core facility for analysis.

Large Batch Purification of MBD2 Constructs

Batch purification of MBD2-G01 and MBD2-G03 constructs was performed as previously described in the Comparison of Recombinant MBD2-MBD from Bacterial and Insect Cells for MBD2-MBD. The only deviation from this protocol was use of a 10,000MWCO concentration column in place of the 3000MWCO column described. Proteins were stored at -80°C after flash freezing in liquid nitrogen.

Dialysis Experiment on MBD2-G03

Same day purified MBD2-G03 protein was used for this experiment. 500µg of protein (approximately 10µL) was added to 490uL of high salt buffer (250mM NaCl, 20mM Tris pH 7.4, 1mM TCEP). Five 100µL aliquots were placed into separate 3000MWCO dialysis cassettes (Thermo Scientific-Pierce Net) after they were pre-soaked in buffer for 5 minutes. Protein was dialyzed overnight at 4°C with gentle mixing by stir bar in 500mL of one of 5 buffers (NaCl at 50mM, 100mM, 150mM, 200mM or 250mM) with 20mM Tris pH 7.4 and 1mM TCEP. Samples were removed the next morning and centrifuged at 17,000xg in a microcentifuge at 4°C. Supernatant was removed and mixed with x4 LDS loading buffer and loaded onto an SDS-PAGE gel (Invitrogen). Gels were visualized by SimplyBlue® staining.

Crystallization of native MBD2-MBD with methylated DNA

MBD2-MBD Crystallization

MBD2-MBD was purified as previously described and concentrated to ~20mg/mL (2mM). Initial crystallization conditions were identified using Sparse Matrix Screening with commercially available screens (Crystal Screen I (Hampton), Crystal Screen II (Hampton), SaltRx (Hampton), PEGs Suite Screen (Qiagen) and Protein Complex Suite (Qiagen)). Hanging drop vapor diffusion experiments were assembled by

mixing 1mM protein (50mM NaCl, 1mM TCEP, 10mM Tris pH 7.4) in a 2:1 ratio with well solution using a Mosquito® liquid handling robot (TTP Labtech) in 96 clear well plates (Corning) with plate covers (Hampton).

MBD2-MBD with Methylated DNA Crystallization

Three protein:DNA complexes, differing in the length of the DNA duplex, were screened for crystallization. Double stranded DNA sequences of 12bp, 14bp and 20bp length and either a T or A single base overhang at either end to assist in crystal lattice formation (Table 2.4). To form complex for crystallization, the DNA was resuspended at 4mM in ddI water and then annealed by heating to 95°C for 3 minutes and then cooling by 1°C per 2 minutes to a final temperature of 25°C. Annealed DNA duplex was mixed in a 2:1 molar ratio to 1mM protein (50mM NaCl, 1mM TCEP, 10mM Tris pH 7.4) and the complex was isolated on an analytical gel filtration complex (analytical Superdex G75). Crystallization conditions were then determined using sparse matrix screening as described previously with the addition of the Natrix Screen (Hampton),

Name	Sequence
20bp_MBD2_MBD_Crystal_F	TCTGGAAmeCGGAATTCTTCTA
20bp_MBD2_MBD_Crystal_R	ATAGAAGAATTCmeCGTTCCAG
14bp_MBD2_MBD_Crystal_F	CATCGTmeCGTACGAC
14bp_MBD2_MBD_Crystal_R	GGTCGTAmeCGACGAT
12bp_MBD2_MBD_Crystal_F	CATCGTmeCGTACG
12bp_MBD2_MBD_Crystal_R	GCGTAmeCGACGAT
14bp_MBD2_MBD_BrdU_Crystal_F	CAbrUCGTmeCGbrUACGAC

Table 2.4: DNA Sequences used for crystal condition screening.

All oligo pairs were annealed prior to screening with MBD2-MBD.

specific for screening DNA containing complexes.

Crystals of MBD2-MBD and 14bp DNA were obtained in 1.5M sodium citrate pH 6.5 and further optimized by grid-screening in 24 well hanging drop plates. Crystals were obtained by mixing 1μL of complex with 1μL of well solution containing citrate between 0.8-2.0M at 6.0 to 8.0 pH. Trays were stored at 25°C and cuboid crystals approximately 200x200x200μm in size formed in 2-5 days.

Crystals were cryostabilized by direct transfer to 2M Citrate pH 6.5 and flash frozen in liquid nitrogen. Data were collected using a Rigaku FR-E x-ray generator (Copper-Kα=1.54Å) with a Saturn 944+ CCD detector and processed with HKL2000(*104*). With two collaborators from the Leahy lab we attempted molecular replacement phasing with Phaser(*105*). A series of heavy metal soaks were performed and data was processed using Shelx(*106*), Solve(*107*) and Resolve(*108*) to identify isomorphous or anomalous signal; however, ultimately, no heavy metal soaks were successful in phasing the protein:DNA complex.

MBD2-MBD Selenomethionine Crystallization

Single point mutations (L193M, R195M, G198M or T200M) were cloned into MBD2-MBD using Quick Change Lightening kit (Agilent Technologies) to aid in phase determination. Selenomethionine substituted protein was expressed and purified as wild-type with the following exceptions. An overnight-culture of BL21(DE3) cells

transformed with MBD2-MBD variants were used to inoculate 1L of minimal media from the M9 SeMET High-yield Media Kit (Medicillon). Cells were grown at 37°C until an OD_{600} of 1.0 and induced as described previously. Upon induction, an amino acid mixture and seleno-methionine slurry were added per manufacturer's instructions to inhibit methionine synthesis and promote selenomethionine incorporation. Cells were grown for 16 hours at 16-20°C and harvested. Protein purification proceeded as described previously except 1mM TCEP was added to all purification buffers.

Complexes between selenomethionine incorporated MBD2-MBD-R195M crystals were obtained mixing 1mM protein (100mM NaCl, 10mM Tris pH 7.4, 1mM TCEP) with 1mM 14bp methylated DNA under conditions 28%-32% PEG8000, 0.2M NaOAc, 0.1M Bis Tris Propane pH 6.5 in hanging drop vapor diffusion 24 well plates. Crystals appeared approximately 2-3 days after trays were assembled and stored at 25°C and were frozen with cryopreservative consisting of 42% PEG8000 or 35% PEG8000/15% Glycerol, 2mM MgCl2, 1mM TCEP, 0.2M NaOAc, 0.1M Bis Tris Phosphate pH 6.5 and stored in liquid nitrogen. SAD data for this mutant was collected on beam line 23-ID-D-GM/CA at the advanced photon source at Argonne National Laboratory with significant help from Jacqueline McCabe and Professor Dan Leahy.

Complexes between selenomethionine incorporated MBD2-MBD-L193M crystals were obtained mixing 1mM protein (100mM NaCl, 10mM

Tris pH 7.4, 1mM TCEP) with 1mM 14bp methylated DNA under conditions 2M ammonium sulfate and 0.1M Na Acetate trihydrate, pH 4.6 in hanging drop vapor diffusion 24 well plates. Crystals approximately 40x40x40μm in size appeared approximately 3-5 days after trays were assembled and stored at 25°C and were frozen with cryopreservative consisting of the same condition above supplemented with 20% glycerol and stored in liquid nitrogen. SAD data for this mutant were collected on beam line 7-2 at the advanced photon source at Argonne National Laboratory with significant help from Pat Byrnes and Professors Dan Leahy and Scott Bailey.

X-ray Crystallographic Data Collection and Structure Determination

Data were collected at beam line 23-ID-D-GM/CA or 7.2 at the Stanford Synchrotron Radiation Light Source (12658.96 eV = 0.9794Å) and processed with HKL2000(*104*)(R195M protein) or XDS(*109*)(L193M protein). Initial heavy atom sites were identified using ShelxC/D(*106*). Single-wavelength anomalous dispersion (SAD) phases were calculated in SOLVE(*107*). A single-round of density modification in RESOLVE (*108*) generated phases that produced interpretable electron density maps.

2.6 Acknowledgements

I would like to make a very special acknowledgement to Dr. Dan Leahy and my collaborators in his lab. Apart from serving on my committee, Dr. Leahy spent hours of his time discussing and explaining the foreign language that is crystallography to an overzealous and completely naïve graduate student. Without his unending patience through numerous crystal structures, including solving both lysozyme and B-form DNA accidentally, we would never have gotten anywhere near where we are now. I truly felt like a member of his lab every moment I was in the Hunterian building, thank you! And to my collaborators in Dan's lab I leave you these thoughts:

I started as a second year graduate student, wide eyed and bushy tailed, with Dr. Ping Liu as my mentor. Ping's hard work and teachings are best described in a story: she was kind enough to entrust her only daughter to my care while she ran my experiments in lab. It was probably mostly because she didn't want to get pink eye (which her daughter had at the time) but I really think it was because she didn't trust me to do just about anything in the lab; a wise decision in my opinion. Through the course of a year I slowly gained Ping's trust and she taught me all the basics behind setting up trays, shooting crystals, collecting data and crossing your fingers and hoping for good diffraction. On top of laying the groundwork for my understanding of this field and for all the future structure work with this protein we collected a native

MBD2-MBD data set July of 2010. My first data set, I'll never forget the excitement. We did go on to solve one structure together over Thanksgiving of 2010, thank you Mr. Lysozyme!

Jackie McCabe, my lab wife. I can't think of anyone who I enjoy arguing with more, except of course my real wife. Hopefully if Sarah reads this she won't be too jealous. I've never met anyone more meticulous and unrelenting than Jackie. No matter how well I thought I had done or thought out an experiment/problem, Jackie was always there to show me where the MBD2 herpes project could improve and why what I did wasn't great, or oftentimes even above average. For her countless hours shooting xtals, collecting and ultimately processing data thank you! We'll always share the zinc based DNA structure we solved together.

And to Pat Byrnes, my current collaborator. While he lacks Ping's height, or Jackie's meticulousness, Pat makes up for in charming looks and thoroughness. Right when the MBD2_MBD project was in a spiraling mess Pat stepped in and valiantly improved the outlook of the project, no small task. For your many mutants, including the one that ultimately will solve this structure, and very late odd hours, thank you!

And finally the dynamic married duo Dr. Jennifer Kavran and Dr. Scott Bailey. I thought it was very fitting that at the end of the day Scott is the one that shot the seleno-met xtal that provided us with our electron density map and Jennifer will be the one that solves it (or holds

someone's hand doing so). While most of the time they looked upon me with pity because of my lack of crystallography charm (and certainly theory) they spent hours of uncompensated time providing guidance and expertise. I will probably never get a chance to repay them for their efforts, so until the next fantasy football game or dinner out thank you! I will certainly be wearing your gift at my defense. And for the record, Ping told me to put that lysate directly on the column…

This work was supported by funding from the Prostate Cancer Foundation, the Patrick C. Walsh Prostate Cancer Research Fund, Department of Defense Prostate Cancer Research Program Pre-doctoral training grant (W81XWH-11-1-0618; to NW), and NIH/NCI grants CA70196 and CA58236.

Chapter 3

Development of a High Throughput Screening Assay for Discovery of Small Molecule Inhibitors of Methyl-CpG Binding Domain Protein 2

3.1 Abstract

Methylated DNA binding proteins such as Methyl-CpG Binding Domain Protein 2 (MBD2) can transduce DNA methylation alterations into a repressive signal by recruiting transcriptional co-repressor complexes. Interfering with MBD2 could lead to re-activation of tumor suppressor genes and therefore represents an attractive strategy for epigenetic therapy. We developed and compared fluorescence polarization (FP) and time resolved fluorescence resonance energy transfer (TR-FRET)-based high-throughput screening (HTS) assays to identify small molecule inhibitors of the interaction between the methyl binding domain of MBD2 (MBD2-MBD) and methylated DNA. The TR-FRET assay (Z' factor = 0.58) emerged as a superior screening strategy compared to FP (Z' factor = 0.08) when evaluated in a HTS 384 well plate format. Using TR-FRET we screened the Sigma® LOPAC library for MBD2-MBD inhibitors and identified 4 compounds that also validated in a dose response series. This included two known DNA intercalators (mitoxantrone and idarubicin) amongst two other inhibitory compounds (NF449, and aurintricarboxylic acid). All four compounds also inhibited the binding of SP-1, a transcription factor with a GC-rich binding sequence, to a methylated oligonucleotide demonstrating that the activity was nonspecific. Our results provide proof-of-principle for using TR-FRET-based HTS to identify small molecule inhibitors of MBD2 and other DNA-protein interactions(*110*).

3.2 Introduction

Epigenetic silencing of tumor suppressor genes via DNA hypermethylation has been established as a common hallmark of oncogenesis(*111*) . The methylation of CpG dinucleotides, particularly at gene promoters and regulatory regions, has been shown to induce epigenetic gene silencing via the recruitment of methyl-binding domain (MBD) proteins such as MBD2, Methyl-CpG Binding Domain Protein 1 (MBD1) and Methyl-CpG Binding Protein 2 (MeCP2) and their associated chromatin remodeling/co-repressor complexes such as Mi2-NuRD(*112*). These complexes are capable of remodeling the local chromatin and preventing the transcriptional machinery from gaining access to DNA, leading to transcriptional repression(*113*). Recent pharmacological interventions for reversal of epigenetic gene silencing in cancer have focused on inhibiting DNA methyltransferases (DNMTs), which establish the DNA methylation marks, and histone deacetylases (HDACs), which are part of transcriptional repressive complexes that signal chromatin compaction via removal of acetylation modifications on histone tails(*114*). Importantly, such efforts have led to FDA approval of two DNMT inhibitors and two HDAC inhibitors for myelodysplastic syndrome and cutaneous T-cell lymphoma respectively(*115*).

However, there has been little effort in developing inhibitors of the methyl binding domain class of proteins, despite the findings from several recent reports credentialing these proteins, particularly the

MBD2 protein, as anti-cancer drug development targets(*116*). At the molecular level, RNAi-mediated depletion of MBD2 led to re-expression of epigenetically silenced tumor suppressor genes with promoter CpG methylation(*117*). *In vivo,* genetic disruption of *Mbd2* in *Apc*^{min} mice, which are prone to developing dozens of intestinal tumors within 4 – 6 months of age(*118*), led to remarkably reduced tumor formation and increased survival(*67*). Interestingly, while mice carrying homozygous disruption of *Dnmt1* alleles show embryonic lethality, mice with homozygous *Mbd2* disruption have a normal life span, size and reproductive potential, suggesting a favorable toxicity profile for targeting MBD2. Taken together, these observations suggest that MBD2 has potential as an anti-cancer drug development target(*116*).

Development of MBD2 antagonists as molecular probes of epigenetic mechanisms and as anti-cancer epigenetic drugs would be greatly aided by the availability of a suitable high-throughput screening assay. Here we describe the development of a modified time-resolved fluorescence resonance energy transfer (TR-FRET)(*119*) assay for measuring MBD2-MBD binding to methylated DNA (Figure 3.1). TR-FRET utilizes the long-lived fluorescence of lanthanide metals to monitor fluorescence resonance energy transfer after a time delay, when auto fluorescent signal has decayed significantly. This translates into a robust signal-to-noise ratio when measuring the binding of two ligands. The TR-FRET assay was highly amenable to high-throughput screening of small

molecule libraries and showed significantly superior performance compared to a fluorescence polarization(*120*) based assay format. We used this TR-FRET screening approach in a pilot screen of 1,280 highly studied compounds, identifying small molecules capable of inhibiting MBD2-MBD binding to methylated DNA.

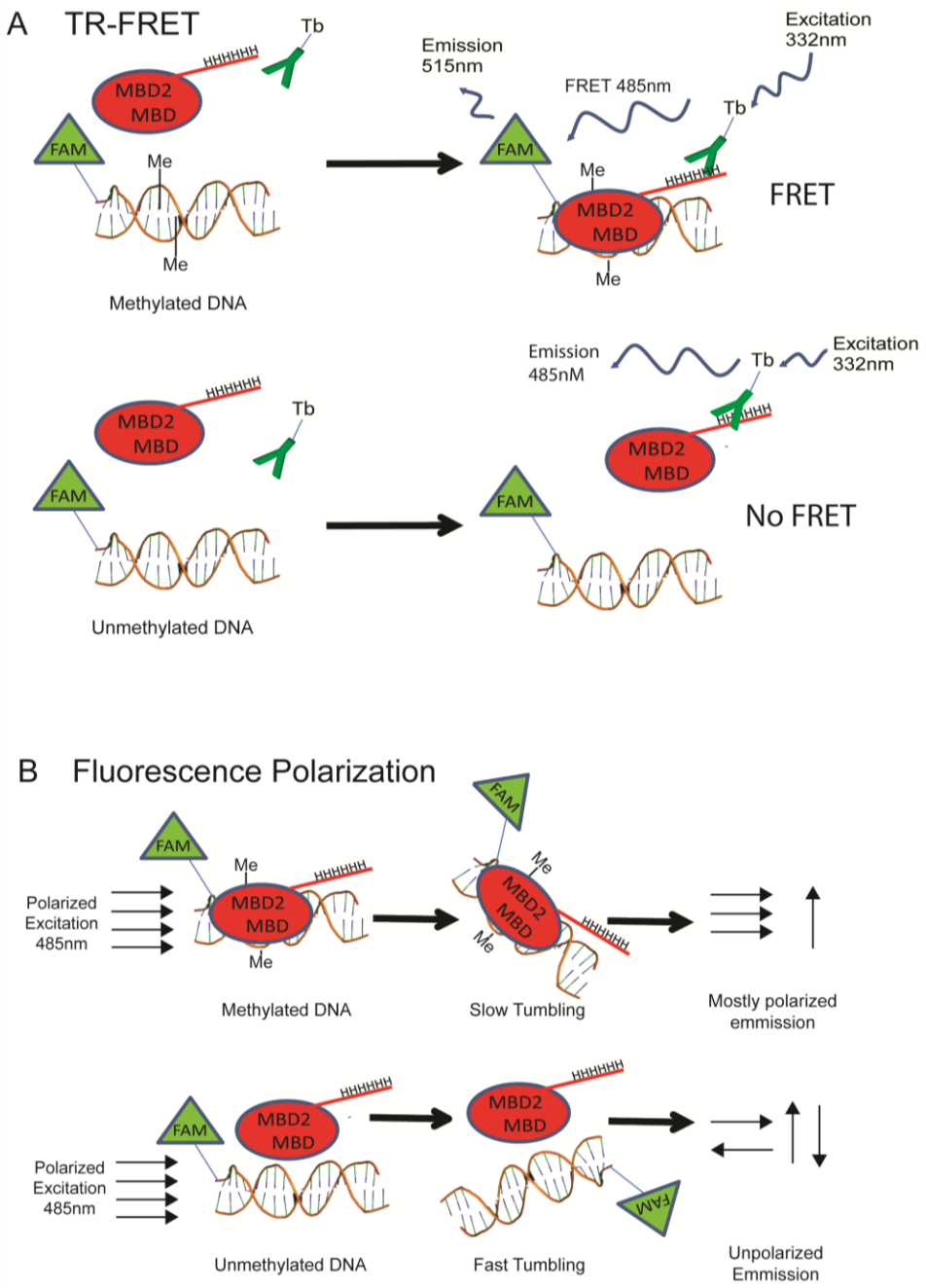


Figure 3.18: Overview of TR-FRET and Fluorescence Polarization MBD2-MBD DNA-binding assays.

(A) TR-FRET overview: MBD2-MBD protein containing a hexa-histidine tag is mixed with FAM-labeled DNA and terbium-labeled anti-penta-His antibody (Tb-Ab). The MBD2-MBD-Tb-Ab-bound complex is excited with a pulse of 332nm laser light and emission is monitored at 485nm and 515nm (result of FRET) after a 50 µsec delay. The ratio of the 515nm and 485nm emission intensity provides a measure of the extent of binding. **(B)** Fluorescence polarization assay overview: MBD2-MBD is incubated with FAM-labeled DNA. The reaction is excited with plane-polarized light, and the extent of polarization of the emitted light is measured using parallel and perpendicular polarization emission filters.

3.3 Results

TR FRET and Fluorescence Polarization Assay Validation

We developed both TR-FRET and FP based assays to measure the binding of MBD2-MBD to methylated DNA, and evaluated each of these assays for use in high-throughput screening for inhibitors capable of disrupting this binding (Figure 3.2). Both assays showed strong preferential binding of MBD2-MBD to methylated DNA (EC_{50} of 86nM and 42.7nM, respectively) compared to unmethylated DNA (EC_{50} >> 1μM) as expected (Figure 3.2A,B) (*63*). Using these experiments, we determined the optimal concentrations of MBD2-MBD, labeled-oligo, and other assay components for each assay (see methods). Using a series of four controls (n = 32 replicates of each), we evaluated the performance of both assays for use in high throughput chemical compound screening, as determined by calculating the Z' factor (Figure 3.2C,D). The TR-FRET assay showed significantly better signal-to-noise ratios (Z' factor = 0.59) compared to that of the FP assay (Z' Factor = 0.08). As expected, the negative control (unlabeled unmethylated oligonucleotide) did not significantly disrupt binding of MBD2-MBD in either assay. Additionally, both assays measured significant DNA binding inhibition by a competitive inhibitor positive control (excess unlabeled methylated oligonucleotide). The TR FRET assay was therefore selected for further validation in a high-throughput screen.

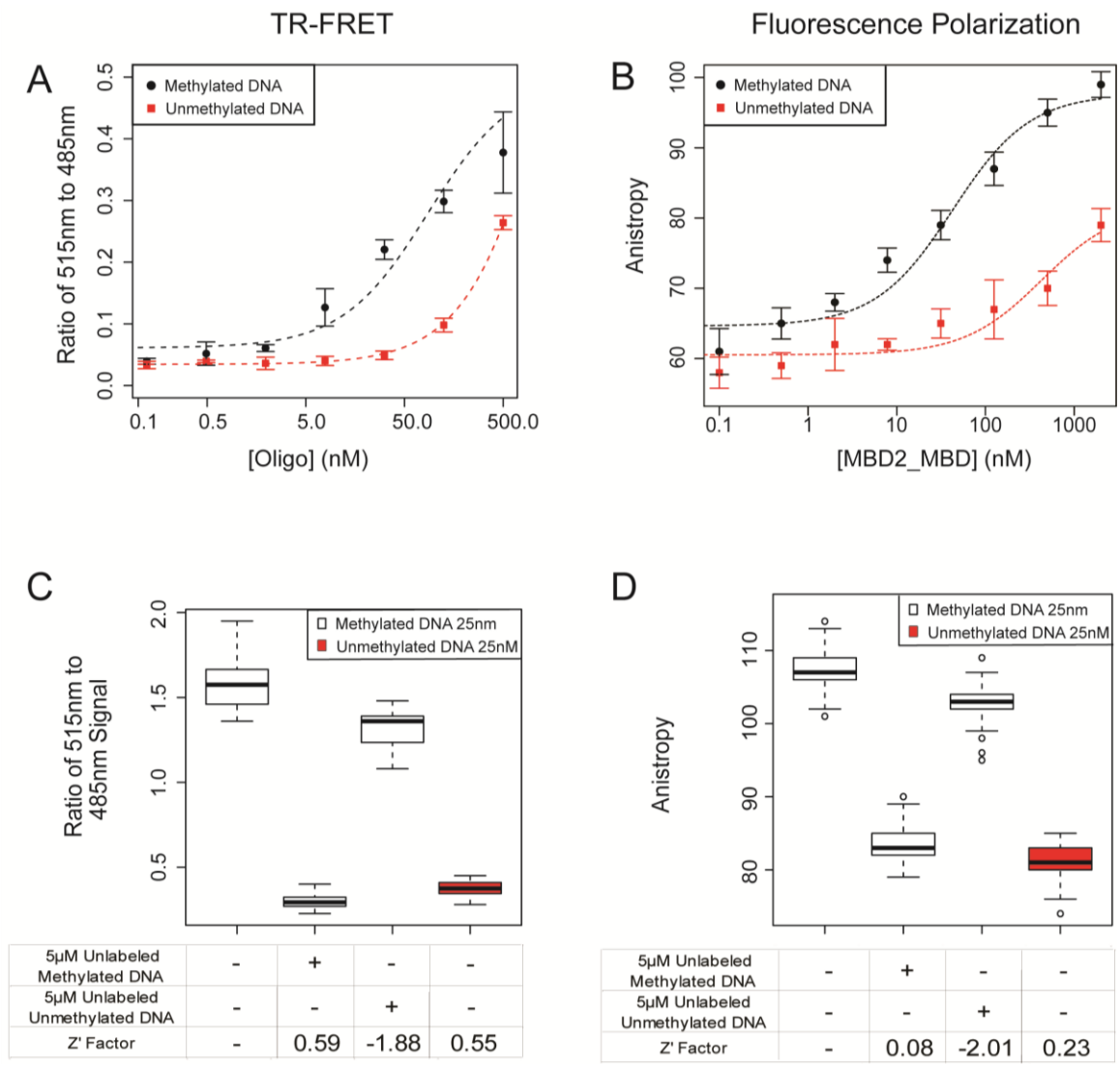


Figure 3.2: Performance of TR-FRET and FP MBD2-MBD DNA binding assays.

(A,B) The binding curves of MBD2-MBD to methylated and unmethylated DNA substrates was measured using the TR-FRET (A), and FP (B) assays. **(C,D)** Performance of the TR-FRET (C) and FP (D) assays with control treatments in 384 well format using 25nM labeled substrate oligonucleotides. Lane 1 (DMSO vehicle control) and Lane 3 (excess unlabeled unmethylated DNA) are negative inhibitor controls; Lane 2 is a positive inhibitor control, consisting of excess fluorophore free methylated DNA. Lane 4 is an assay negative control, showing lack of binding with FAM labeled unmethylated DNA. Z' factors are calculated relative to the DMSO vehicle control (Lane 1).

Pilot screen of LOPAC$^{1280\ TM}$ compounds for inhibition of MBD2-MBD binding to methylated DNA

To assay inhibitors of MBD2_MB2 binding a methylated oligonucleotide, we used the 1280-compound Library of Pharmacologically Active Compounds (LOPAC™) (Sigma Aldrich®). This library consists of a variety of drug and drug-like compounds with well-studied mechanisms, as well as molecules from larger chemical compound libraries. Test compounds from the LOPAC™ were added to our optimized TR-FRET assay at a final concentration of 20μM. The compounds were assayed across four 384 well plates with each plate containing positive and negative controls (Figure 3.3A). The TR-FRET signal from most compounds followed a normal distribution centered around the median of the negative control. Z and Z' factors were calculated for each plate (Figure 3.3B); Z' factors for each plate were 0.69, 0.73, 0.73 and 0.55 respectively. As expected, Z factor scores for each plate were lower than their respective Z' score but all were above 0.5. The LOPAC™ screen replicated the sensitivity of the initial validation and also Identified ten compounds that significantly inhibited MBD2-MBD binding to methylated DNA after correction for multiple hypothesis testing (Figure 3.3A, open circles, Table 1).

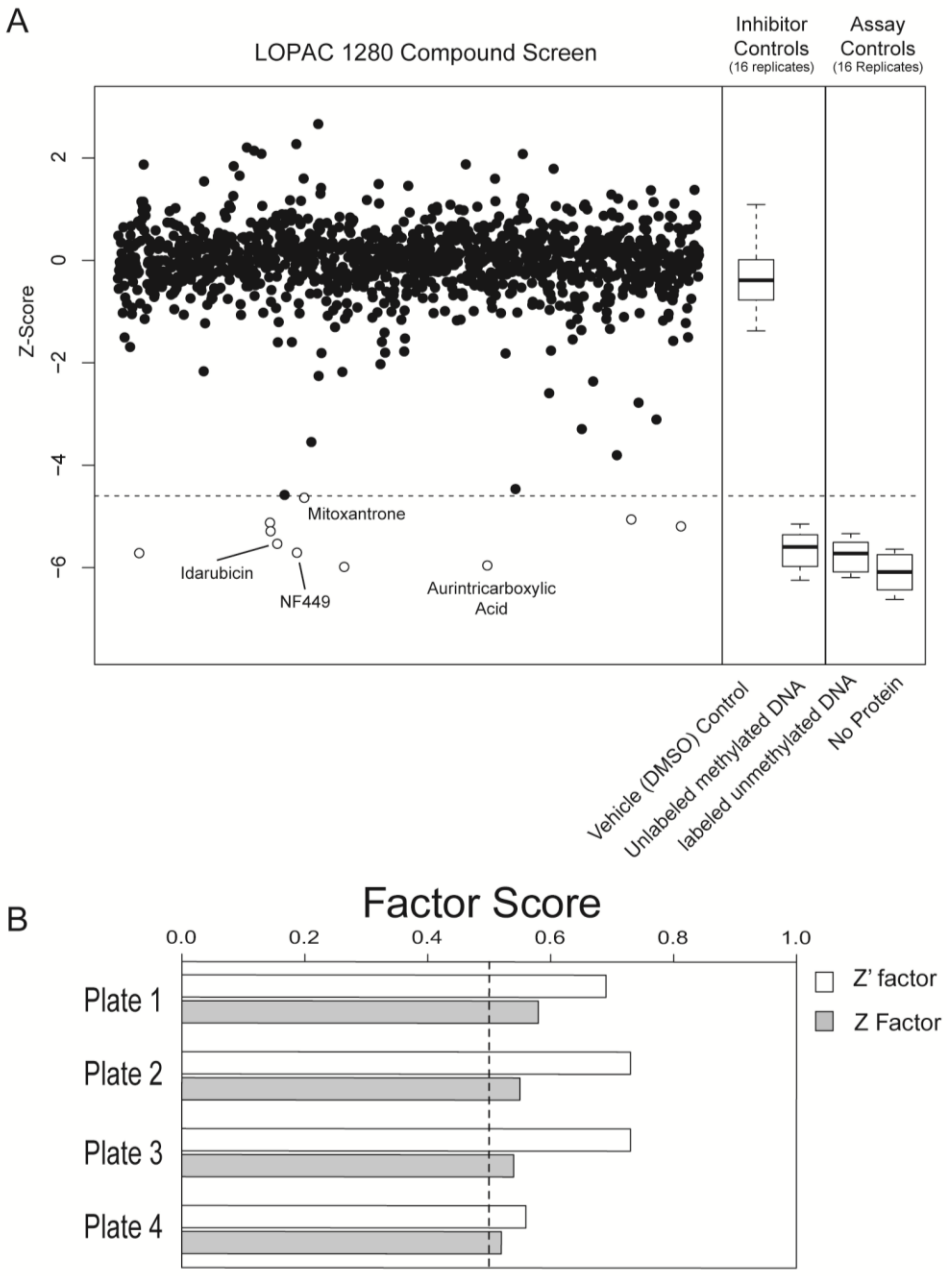


Figure 3.3: Screening of LOPAC$^{1280™}$ library with the MBD2-MBD TR FRET assay.

(A) TR FRET ratios for each compound were normalized to the plate median and plate standard deviation to obtain a standardized Z-score that was plotted for each compound. Open circles represent compounds with significant binding inhibition. The dotted line represents the Z-score cutoff corresponding to a Bonferroni adjusted p-value < 0.05. Each of the four plates contained 16 replicates of four controls, including two inhibitor controls and two assay controls (see box plots). Those primary hit compounds (open circles) that were confirmed through dose-response analysis (see Figure 3.4) are labeled. **(B)** Z' and Z factors for the screen data were calculated for each plate according to equation (1) as described in Materials and Methods. In all cases Z' was greater than Z and all values were above 0.5 (indicated by the dotted line).

Compound*	Primary Screen P-Value	MBD2-MBD Secondary Screen IC50	SP-1 Counterscreen IC50	Description
6-Hydroxy-DL-DOPA	2.16E-07	>50μM	Not Assayed	Precursor of the catecholaminergic neurotoxin, 6-hydroxydopamine; converted to 6-hydroxydopamine by L-aromatic amino acid decarboxylase.[1]
Aurintricarboxylic acid	2.56E-07	2.7nM	<1nm	DNA topoisomerase II inhibitor. Polymerizes in water.[1]
Dephostatin	5.50E-07	>50μM	Not Assayed	CD45 protein tyrosine kinase inhibitor.[1]
NF449 octasodium salt	7.08E-07	290nM	24.7nM	Potent G protein antagonist, selective for Gs-alpha[1]. P2X1 selective protein inhibitor.[2]
Idarubicin	7.90E-07	196nM	580nM	Antineoplastic[1]. DNA intercalator.
Reactive Blue 2	1.67E-06	Not Assayed	Not Assayed	P2Y receptor antagonist; most potent antagonist for ATP-activated channels.[1]
Me-3,4-dephostatin	1.74E-06	>50μM	Not Assayed	Selective inhibitor of protein tyrosine phosphatase 1B and SHPTP-1.[1]
Carbetapentane citrate	5.19E-06	>50μM	Not Assayed	Opioid receptor ligand with high affinity toward sigma1 sites; antitussive.[1]
(2S,1 S,2 S)-2-(carboxycyclopropyl) glycine L-CCG-I	5.69E-06	>50μM	Not Assayed	Potent group II metabotropic glutamate receptor agonist.[1]
Mitoxantrone	1.72E-05	94nM	314nM	DNA synthesis inhibitor[1]. DNA intercalator.

(1) Sigma Aldrich (http://www.sigmaaldrich.com/chemistry/drug-discovery/validation-libraries/lopac1280-navigator.html)
(2) El-Ajouz, S., et al, Molecular basis of selective antagonism of the P2X1 receptor for ATP by NF449 and suramin: contribution of basic amino acids in the cysteine-rich loop. Br J Pharmacol, 2012. 165(2): p. 390-400. (Ref. 21).

Table 3.1: Significant hits from the Sigma® LOPAC$^{1280™}$ screen

Dose response of hits for inhibition of MBD2-MBD binding to methylated DNA

To validate the "hits" observed on the LOPAC™ screen, we assayed each compound in a dose-response (Figure 3.4, Table 1). A suramin analog, Reactive blue 2, emerged as a "hit" from the primary screen but was excluded from the secondary analysis because its coloration and light absorption could interfere with the assay. Of the remaining 9 compounds, four compounds (mitoxantrone, idarubicin, aurintricarboxylic acid, and NF449) showed a sigmoidal dose-response for inhibition of MBD2-MBD binding to methylated DNA, while five compounds (6-Hydroxy-DL-DOPA, Carbetapentane citrate, Dephostatin, Me-3,4-dephostatin, and L-CCG-I) failed to show appreciable inhibition in the dose range tested (Figure 3.4, Table 1). Mitoxantrone and idarubicin, are known DNA intercalators (*121, 122*), and aurintricarboxylic acid is known to polymerize in water and can inhibit many macromolecular protein-nucleic acid interactions (*123*). These three compounds would be expected to nonspecifically inhibit DNA-protein interactions. Interestingly, a suramin analog, NF449, known to inhibit P2X1 receptors with picomolar potency and high selectivity compared to P2X2 receptors (*124*), also emerged as a dose-response validated inhibitor of the MBD2-MBD interaction with methylated DNA, showing an IC_{50} of 290nM.

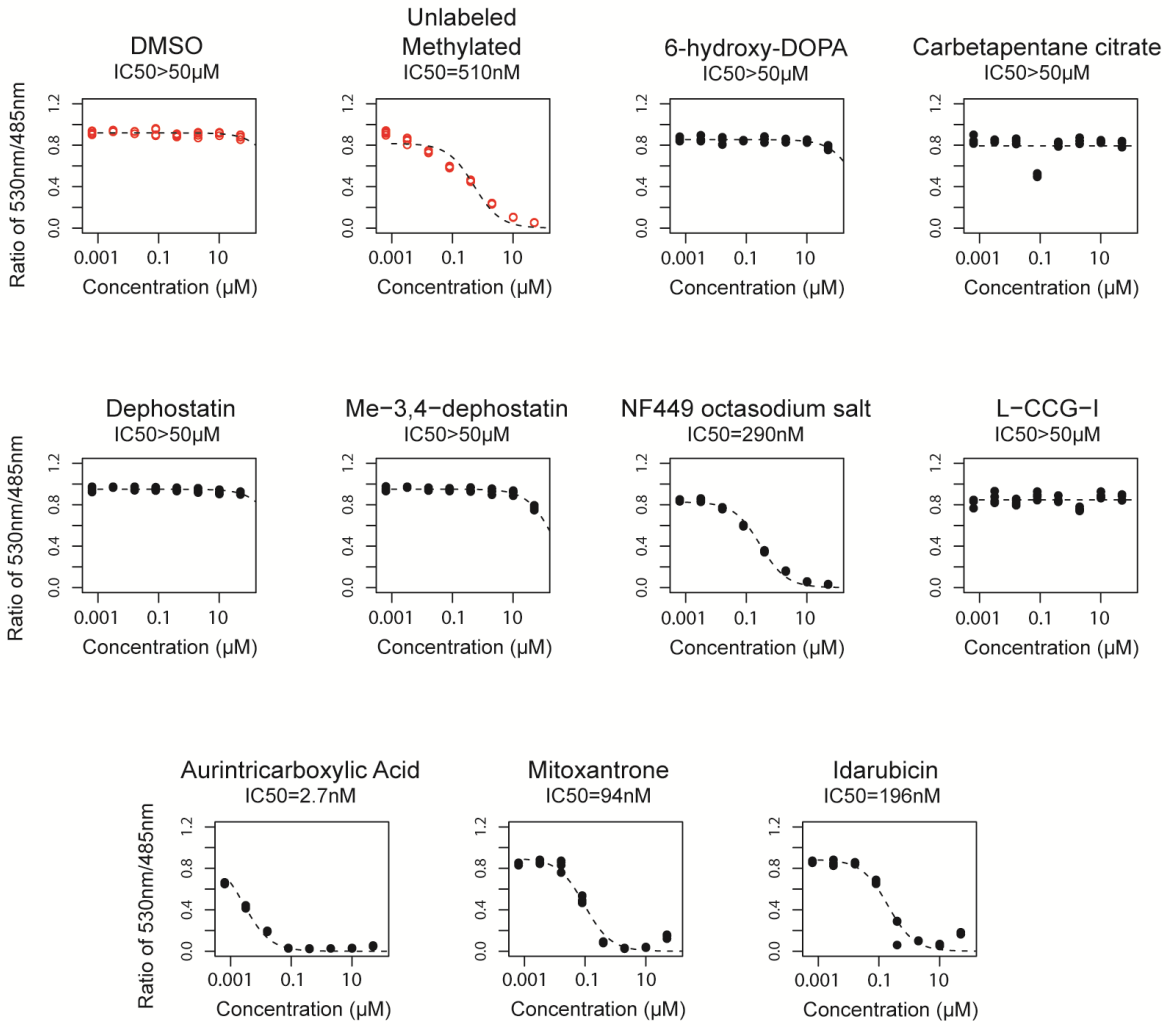


Figure 3.4: Dose response curves with significant hits

The 9 hit compounds from the primary screen were tested with the TR-FRET assay using a 4-fold dose response series from 1.2nM to 20μM (4 replicates each). Controls are shown as red open circles. Four of the 9 compounds showed a potency for inhibition of MBD2-MBD binding to DNA within the dose range tested. IC50 values for each compound are listed below the title for each dose response.

Counter screen with the DNA Transcription factor SP-1

To assess the specificity of the four verified hit compounds for inhibiting binding of MBD2-MBD to methylated DNA, we developed a counter screen (Figure 3.5A) using the transcription factor SP-1 to identify those hits that would also inhibit binding of an unrelated protein to DNA or some other components of the assay. The transcription factor SP1 has been shown to localize to GC rich regions in the human genome, including promoter CpG islands where MBD2 can be found (*117*). However, SP-1 has no known specificity for methylated DNA, which we verified (EC_{50} = 22nM for both methylated and unmethylated DNA (Figure 3.5B) (*125*). We then tested MBD2-MBD inhibitors from the LOPAC™ for cross-inhibition of SP-1. All four hit compounds inhibited SP-1 in a dose-response manner (Figure 3.5C), suggesting that these compounds are unlikely to be selective inhibitors of MBD2-MBD binding to DNA, but may have activity in inhibiting DNA-protein interactions more generally.

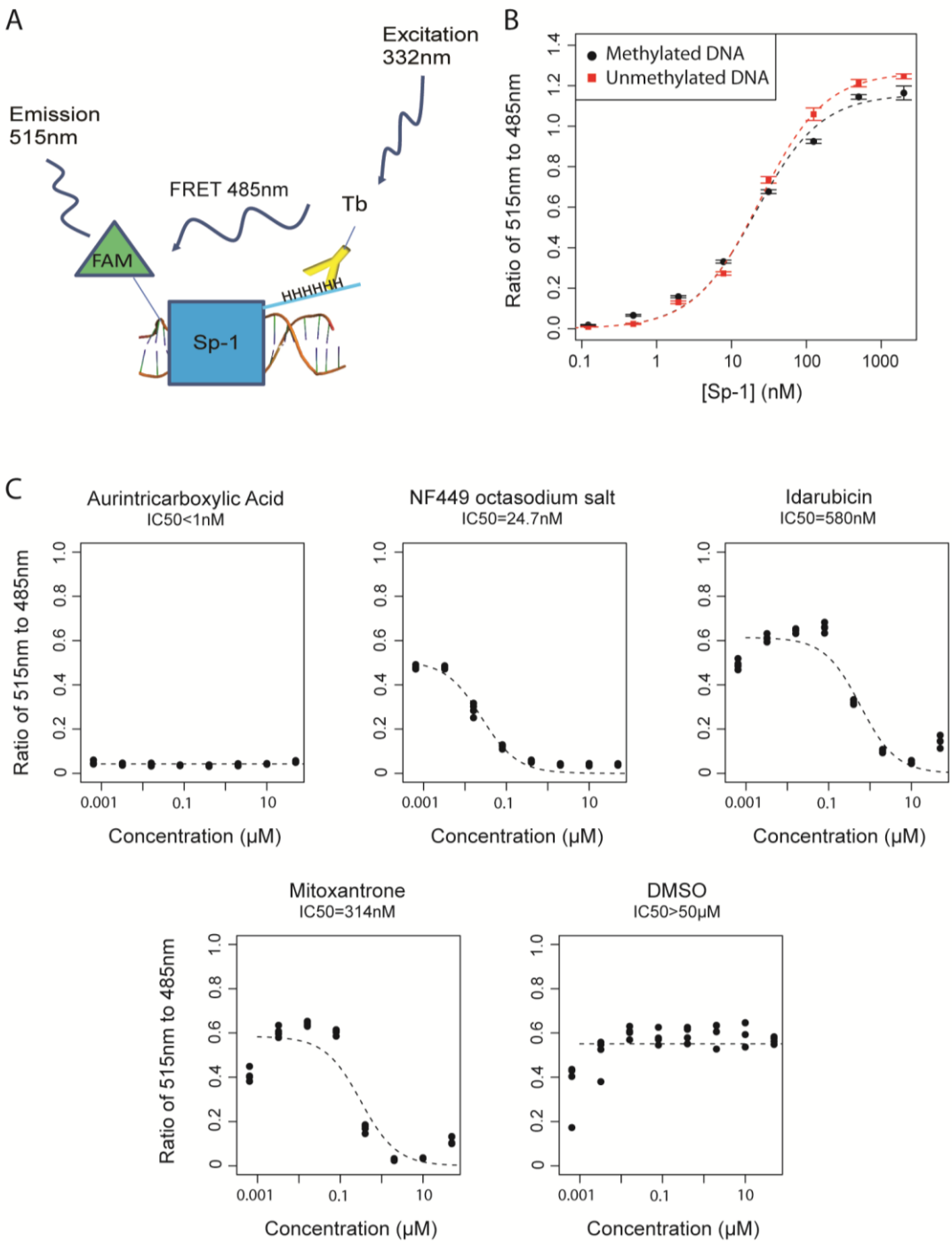


Figure 3.5: Counter-screen with DNA transcription factor SP1 to assess specificity of identified hit compounds.

(A) Overview of TR-FRET assay for measurement of SP-1 binding to DNA. **(B)** Transcription factor SP-1 binding to methylated and unmethylated oligonucleotides in the TR FRET assay. There appears to be no dependence of the methylation status on protein-DNA binding. (C) All 4 compounds showing a dose-responsive inhibition of MBD2-DNA binding were also capable of inhibiting SP1-DNA binding in a dose-dependent manner.

3.4 Discussion

We have developed a robust, high throughput, TR-FRET based assay for screening small molecule libraries for inhibitors of MBD2_MBD protein binding methylated DNA. Compared to a fluorescence polarization based approach, our TR-FRET assay provides a better signal-to-noise ratio, resulting in improved Z' factors under optimized screening conditions. We used this assay to analyze the 1280 compound LOPAC™ library to identify small molecules capable of inhibiting the interaction of MBD2-MBD with methylated DNA. Of the initial 9 hits, four showed a robust sigmoidal inhibitory dose-response. Three of these four compounds, mitoxantrone, idarubicin, and aurintricarboxylic acid, would be expected to inhibit MBD2-MBD binding to methylated DNA given that each are known to inhibit multiple DNA-protein interactions. This was confirmed by our counter-screen which showed that these compounds also prevented binding of SP1 to DNA with similar potency. However, the ability of NF449 to inhibit MBD2-MBD binding to methylated DNA was not anticipated.

NF449 is a suramin analog which can inhibit the P2X1 receptor with sub-nanomolar potency (*126, 127*). The P2X family form homo or heterotrimeric complexes on the surface of many cell types, forming ATP-gated cation channels (*128*) that have been implicated in multiple cellular processes (*129*). The mechanism of inhibition features ionic interactions between the 8 sulfonic acid groups on NF449 with positively

charged basic amino acids at the base of a cysteine-rich loop on P2X family members (*127*). The selectivity of NF449 to P2X1 compared to other P2X and P2Y members is thought to arise through a second favorable contact between the compound and lysine 138 near the ATP binding pocket (*127*). NF449 has also been shown to inhibit other proteins, including the G protein alpha subunit (*130*) and FGFR3 (*131*). In each case, the number and orientation of the negatively charged sulfonic acid groups on NF449 were thought to be critical for inhibition. We speculate that NF449 may similarly inhibit MBD2: the negatively charged sulfonic acids on NF449 may disrupt the binding of positively charged basic residues on the MBD2 surface with the negatively charged phosphate backbone of DNA (*99*). A similar mechanism may also be responsible for the observed disruption of SP1 binding to DNA by NF449 in our counter screen. Additionally, the previous work examining the inhibition of FGFR3 by NF449 suggested that the compound may be capable of interacting with the intracellular tyrosine kinase domain of FGFR3 (*131*). NF449 may therefore be available in the intracellular space for activity. If this proposed intracellular bioavailability is confirmed, then NF449, or its derivatives, may have promise as pharmacological probes for studying MBD2 function in cells. Additionally, in future work, it may be possible to find NF449 derivatives that are capable of selectively inhibiting MBD2 and not other DNA binding proteins,

analogous to the selectivity of NF449 for P2X1 compared to other P2X receptors (*127*).

These data demonstrate the utility of our TR-FRET based MBD2 methylated-DNA binding assay and coupled counter screen to identify and test the specificity and activity of small molecules. This TR-FRET based primary screening and counter-screening assay should be easily adaptable to more extensive small molecule libraries in order to discover novel, specific inhibitors of the MBD2-methylated-DNA interaction.

3.5 Methods

MBD2-MBD Production

A codon optimized sequence for the MBD2-MBD polypeptide was synthesized and cloned into the pGSE6 vector (Genscript USA Inc) for expression in bacteria as a C-terminal hexa-histidine tagged fusion protein. Briefly, BL21 DE3 cells (Agilent Technologies) were transformed with this construct, allowed to grow to an OD_{600} of 1.0, and were induced with 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG, Corning Cellgro) overnight in a shaking incubator at 220 rpm and 20°C. Bacteria were lysed using a French press in equilibrium buffer containing 300 mM NaCl, 50 mM sodium phosphate pH 8.0, 5 mM imidazole (Sigma Aldrich), and a protease inhibitor cocktail (Roche USA). Lysates were centrifuged at 17,000x*g* for 30 minutes and then incubated with IMAC Nickel NTA beads (Bio-Rad) for 1 hour at 4°C. The beads were washed three times with equilibrium buffer containing 15 mM, 20 mM and 25 mM imidazole, respectively, and affinity purified protein was eluted in equilibrium buffer containing 150mM imidazole. The his-tagged protein was further purified by gel filtration using a Superdex g75 26/300 Column (GE Healthcare Life Sciences) in storage buffer (150mM NaCl, 20mM Tris pH 7.4, 1mM tris(2-carboxyethyl)phosphine (TCEP, P212121 LLC, Toledo, Ohio). Fractions containing the purified MBD2-MBD were concentrated with 3000 MWCO Amicon® Ultra-4 Centrifugal Filter Units (Millipore) and flash frozen with liquid nitrogen and stored at -80°C.

Fluorescence Polarization Assay

Fluorescence polarization binding assays were performed as previously described (*117*). Briefly, MBD2-MBD protein was added in a dilution series to the reaction buffer containing 4% glycerol, 1 mM MgCl2, 0.5 mM EDTA, 0.5 mM DTT, 125 mM NaCl, 10 mM Tris–HCl (pH 7.4), 0.2% Tween-20 up to a total working volume of 20μL. Hairpin-forming oligonucleotides with the sequence 5′-fluorescein (FAM)-ATGCTCGTAGCACTTTTGTGCTACGAGCAT-3′ (unmethylated) or 5′-fluorescein - ATGCTC^{me}GTAGCACTTTTGTGCTAC^{me}GAGCAT-3’ (methylated) were annealed by heating to 95^{o}C and cooling on ice rapidly. They were then added to the binding reaction to a final concentration of 10nM. Binding assays were carried out in 384-well black, flat-bottom plates (Corning) in quadruplicate replicates. The plate was incubated at 4^{o}C for 1 hour with gentle shaking. Fluorescence polarization readings were performed on a Safire^{2} (Tecan) instrument with excitation at 470nm and monitoring emission at 525nm. Anisotropy values were acquired and plotted relative to MBD2-MBD protein concentration and fitted using the open source software R version 2.12.1 (*132*). The optimal conditions of 10nM oligonucleotide and 200nM MBD2 protein were used for the high throughput assay. Positive and negative inhibitor controls consisted of 5μM of hairpin oligonucleotides (either methylated or unmethylated) with the same sequence described above but without a FAM label, and were

analyzed in 32 replicates each. Z' factors were calculated using the equation (1):

$$Z' = 1 - 3\,x\,\frac{SD_m - SD_u}{|\mu_m - \mu_u|} \qquad (1)$$

where SD_m and μ_m are the standard deviation and mean anisotropy values respectively of protein binding to the methylated FAM labeled oligonucleotide, and SD_u and μ_u are the standard deviation and mean of anisotropy values respectively of the protein binding to the unmethylated FAM labeled oligonucleotide (*133*).

Time Resolved Fluorescence Resonance Energy Transfer (TR-FRET)

TR-FRET assays were performed in 384-well black low volume plates (Greiner Bio One) with a total working volume of 20µL. MBD2-MBD protein at 25nM was added to the same buffer as described above for fluorescence polarization. The FAM-labeled oligonucleotides described above were added in a dilution series and LanthaScreen® Elite Tb-anti-His-Tag Terbium-labeled antibody (Life Technologies) was added to a final concentration of 5nM. The plate was incubated at 4°C with gentle shaking for 1 hour. The plates were read on a Safire^2 (Tecan) instrument with excitation at 332nm and emission read at 485nm (to read FAM emission) and 515 nm (to read Terbium FRET emission) after a delay of 50 µsec and a total integrated read time of 400 µsec. The ratio of the 515 nm and 485 nm readings was used to assess the degree of FRET, which is proportional to the amount of total binding. The ratio from a control

sample containing no protein was subtracted from the other data and the binding curves plotted relative to total DNA concentration. Curve fitting was performed using the open source software R version 2.12.1. The optimal conditions were determined to be 30nM oligonucleotide and 25nM MBD2 protein. Positive and negative inhibitor controls consisted of 5μM of oligonucleotide (either methylated or unmethylated) with the same sequence described above but without a FAM label, and were analyzed in 32 replicates. Z' factors were calculated in the same manner listed above.

LOPAC$^{1280\ TM}$ Screen

The 1280 compound Library of Pharmacologically Active Compounds (LOPAC)™ (Sigma-Aldrich) was screened using the TR-FRET assay described above with optimized conditions in 384-well low volume plates (Greiner Bio-One) in a reaction volume of 20μL. A final concentration of 20μM in 10% dimethyl sulfoxide (DMSO) was used for each compound. Four controls were used on each plate with 16 replicates for each control: 1) a 10% DMSO vehicle control, 2) 5μM of unlabeled, methylated oligonucleotide described above as a competitive inhibitor positive control, 3) unmethylated, FAM-labeled oligonucleotide described above as a technical control, and 4) a second technical control in which no MBD2-MBD protein was added. Plates were incubated for 1 hour at 4^{o}C and read using the Safire^{2} (Tecan) and the conditions listed above. The

emission ratios were calculated as described above and were adjusted based on the median plate value for each plate. For each compound i on plate j, the Z-score was calculated as shown in equation (2):

$$Z\ Score = \frac{x_{ij} - X_j}{\sigma_j} \qquad (2)$$

where $x_{i,j}$ is the TR-FRET signal for each compound, X_j is the median plate TR-FRET signal across all compounds on that plate, and sigma is the standard deviation across all compounds on plate j. The two-sided p-value associated with the Z-score was calculated assuming a normal distribution and was subjected to Bonferroni correction. Compounds with Bonferroni adjusted p-value < 0.05 were considered significant hits. The Z' factor was calculated for each plate separately using equation (1). Additionally, a Z-factor was calculated for each plate according to equation (1), except that SD_m and mean μ_m represent the standard deviation and mean ratio values respectively for all compounds on each plate.

All compounds that significantly inhibited protein binding in the LOPAC screen were tested in a dose-response series (1nM to 20μM) using the TR FRET assay. Also included were a negative vehicle control (DMSO only) and a positive inhibitor control using 5μM unlabeled methylated oligonucleotide. Curve fitting was performed using the open source software R version 2.12.1.

Counter screen with the Transcription Factor SP-1

We adapted the TR-FRET assay for assessment of recombinant Hexa-His-tagged SP1 protein binding to the unmethylated hairpin oligonucleotide described above. The assay was performed using 10nM of the DNA binding transcription factor SP1 (Abcam), 100nM FAM-labeled oligonucleotide, and compounds in a dose response from 1nM to 20μM, with the same plate type and conditions described above.

3.6 Acknowledgements

The authors would like to thank the laboratory of Dr. Dan Leahy, Dr. Ping Liu and Jackie McCabe for expertise with protein expression and use of their gel filtration system. The authors would also like to thank the Johns Hopkins School of Medicine High Throughput Biology (HiT) center, in particular Dr. Min Li, Alan Long and Melissa Miller, for assistance with the LOPAC screen and for helpful discussion.

This work was supported by funding from the Prostate Cancer Foundation, the Patrick C. Walsh Prostate Cancer Research Fund, Department of Defense Prostate Cancer Research Program Pre-doctoral training grant (W81XWH-11-1-0618; to NW), and NIH/NCI grants CA70196 and CA58236.

Chapter 4

High Throughput Screen to Discover Small Molecule Inhibitors of the Methyl Binding Domain of MBD2

4.1 Abstract

In the past two decades a significant effort towards discovering small molecule inhibitors of epigenetic proteins has occurred. Despite mounting evidence that epigenetic alterations are not only stably selected but also drivers of many cancers only 4 drugs have received FDA approval. These medications target either the DNA 5-methylcytosine writer protein DNMT1 or the histone eraser protein family known as HDAC's. In this work we use a high throughput screening approach to find small molecule inhibitors of the DNA methylation reader protein MBD2, a novel epigenetic target.

In collaboration with the Scripps Institute® we utilized a TR FRET based assay (chapter 3) to screen the NIH's MLPCN small molecule library consisting of 370,276 compounds. We found 1,149 compounds (0.31%) able to significantly disrupt MBD2-MBD from binding methylated DNA (meDNA). The assay performed as expected with a Z'=0.84±0.01 and Z=0.79±0.10. Of the primary screen active compounds 271 validated in triplicates at a single dose. We employed a counter screen using another methyl cytosine reader protein, UHRF1, also in a TR FRET based assay, to find nonspecific DNA-protein disruptors. Comparing the counter screen and secondary screen we classified 48 compounds as both significant binding disruptors and as being MBD2 specific. 46 of the active compounds were subjected to a dose response

with our UHRF1-SRA and MBD2-MBD TR FRET assays, identifying 18 as both MBD2-specific and having an $IC_{50} \leq 10\mu M$.

MCF7 and LNCaP cancer cell lines were subjected to 17 of the 18 active compounds found by our screen in both a single and dose response series. This treatment resulted in 6 compounds causing significant re-expression of epigenetically silenced genes. All but one of these small molecules also caused luciferase expression on a plasmid containing a fully methylated GSTP1 promoter driving luciferase.

Biophysical analysis showed that none of the 17 active compounds altered the melting temperature of MBD2-MBD protein to a significant extent in a thermal melt assay. A DNA intercalator competition assay revealed that all but 3 of the original 17 hit compounds significantly displaced Sybr® green from DNA, suggesting the majority of our hits interact with DNA in a methylation and protein independent manner. Only 2 of the compounds were capable of relieving epigenetic repression in our cell based assays. The binding specificity and IC_{50} results were confirmed by isothermal calorimetry experiments. Interestingly NF449, a small molecule identified in the LOPAC$^{1280™}$ screening efforts, bound MBD2 protein in a specific and dose responsive manner by ITC.

We investigated the mechanism of action for the two most biologically active compounds, KCC-120 and KCC-111, by performing ChIP experiments on MCF7 cells. Interestingly, at methylated and unexpressed promoters both compounds caused a shift towards a more

open chromatin state with increases in H3K4 trimethylation and H3K9 acetylation and loss of H3K27 and H3K9 trimethylation. These changes were not observed at unexpressed but unmethylated promoters or highly active promoters. Some loss of MBD2 at methylated promoters was also observed, providing a possible explanation for this observation.

The HTS and follow-up experiments ultimately identified three possible MBD2-MBD interacting compounds: KCC-127 and 131, which were inactive in the ITC, protein melt curve or DNA intercalator assays, and NF449, which shows micromolar affinity to MBD2 *in vitro*. Two DNA interacting compounds, KCC-111 and 120, significantly relieve epigenetic repression and do so by causing chromatin changes and possibly disrupting MBD2 binding. Work is ongoing to determine whether KCC-131 and 127 are targeting MBD2 and if NF449 has biological activity. Through these efforts we've identified lead compounds for interruption of a novel epigenetic target.

4.2 Introduction

It is well established that epigenetics plays an important role in a number of neoplastic malignancies (*134-137*). A significant effort, particularly over the past decade, to find novel small molecule inhibitors of key epigenetic proteins has thus emerged. To date, medications targeting the epigenetic writer protein DNMT1 and eraser HDAC proteins have won FDA approval, all with hematopoietic indications (*19*). Because of these successes the majority of work in epigenetic therapy has been spent investigating novel DNMT and HDAC inhibitors. Other targets have recently entered the field and are of intense interest. For example bromodomain and extra terminal (BET) and bromodomain containing proteins have become extremely popular targets due to recent evidence that their inhibition significantly hinders proliferation and aggressiveness in numerous cancer models(*138*). We will briefly discuss the key epigenetic targets and the efforts towards inhibiting them for therapeutic efficacy.

The first epigenetic protein targeted by small molecules was the DNA methylation writer protein DNMT1, with azacytidine winning FDA approval for all subtypes of myelodysplastic syndrome (MDS) in 2004(*139*). Azacytidine, in three clinical trials run by the Cancer and Leukemia Group B (CALGB), improved life expectancy relative to supportive care, but did not change the overall conversion rate to acute

myeloid leukemia (AML), the most lethal outcome for this disease. Azacytidine is a cytidine mimic and is incorporated into both DNA and RNA, disrupting translation machinery as well as causing hypomethylation and replication fork collapse (*139, 140*). A number of second generation DNMT1 inhibitors are in development (i.e. zebularine) or are already FDA approved (i.e. decitabine) and remove the RNA incorporation seen with azacytidine(*141*). This significantly lowers unwanted side effects and toxicity(*141*). The mechanism of action for this class of inhibitors is complicated. The cell death caused by DNA replication collapse seen in the setting of high levels of azacytidine is easily understood, as most chemotherapeutics act in a similar manner. The epigenetic mechanism of action is a bit less clear. In an ideal setting DNMT1 inhibition relieves DNA methylation based silencing of a tumor suppressor gene, such as p16, allowing it to resume normal function and cause apoptosis or senescence(*142*). While evidence exists that this occurs, increases in reactive oxygen species and the major chromatin structure changes that occur with DNMT1 inhibition may also contribute to the therapeutic effect (*142-144*). The ultimate mechanism behind the pharmacodynamics is likely a combination of these processes, but that remains to be proven. A recent crystal structure of DNMT1 bound to DNA has rekindled interest in this target and a number of groups have begun efforts at structure aided design to find novel non cytidine mimicking

small molecules(*145*). Significant interest remains in finding novel DNMT1 inhibitors with less toxicity and greater efficacy.

The second class of proteins currently targeted in the clinic is the histone deacetylase family. In 2006 the FDA approved vorinostat, a pan-HDAC inhibitor and the first non-DNMT1 epigenetic targeting agent certified for patient use. Vorinostat approval occurred following an open label trial which demonstrated significant delay in progression time in patients with cutaneous T-cell lymphoma (CTCL) pretreated with the medication(*20*). Quality of life scores also significantly improved under this therapy regimen. Since vorinostat's approval, only romidepsin, also for CTCL, has made it past clinical trials as an HDAC inhibitor(*21*). Both of these agents are pan inhibitors, targeting numerous classes of HDAC's. Interestingly, despite targeting different HDAC classes, the mechanism of action for both vorinostat and romidepsin are the same, involving abrogation of a key zinc metal and the enzymatic active site (*142, 146*). Similar to DNMT1 inhibitors, the mechanism behind the therapeutic index is multifaceted: dependent upon up regulation of tumor suppressors, silencing of anti-apoptotic genes, increased levels of reactive oxygen species, and anti-angiogenic factors(*146*). A significant amount of work remains to be done to fully understand the ideal HDAC classes or histone eraser proteins to target, as well as find more potent inhibitors of this class of proteins (*147, 148*).

More recently, a wave of enthusiasm for targeting another epigenetic reader protein has swept through the scientific community. Similar to methylated DNA, acetyl histone modifications are read by a large family of reader proteins that assist with enforcing the chromatin state. The domain responsible for identifying this epigenetic mark and mediating downstream gene expression is known as the bromodomain(*149*). The human genome contains 42 genes that encode for at least one of 61 bromodomains(*150*). These proteins often hold additional classes of histone reading domains, such as HAT or PHD domains(*151*). Bromodomains themselves have no enzymatic activity and rely on recruiting complex members or positioning other domains in the same protein to promote gene expression(*150*). Much of the excitement behind this class of proteins stems from the specificity provided by their structures(*152*). Though the binding pocket of all bromodomains is hydrophobic and well conserved, the entry into the pocket contains numerous unique residues which make it possible to design protein specific inhibitors (*153-155*). Recently, two groups successfully targeted BET containing bromodomain family members, BRD4 and BET151, with nanomolar affinity(*156, 157*). Interestingly, BRD4 inhibition shows significant anti-proliferative and anti-anchorage independent growth effects in a number of different cancer cell models *in vitro* and *in vivo* (*158, 159*). Clearly the bar is set high for these novel inhibitors as they enter more complicated animal trials and early clinical trials.

While the brief discussion here hardly covers the complete scope of epigenetic proteins under investigation as targets for therapy it is clear that a significant effort is underway. A large body of work on targeting the writer DNMT1, histone reader BET/bromodomain and histone eraser HDAC exists. Little of these efforts, however, have focused on DNA methylation readers. In the previous two chapters we describe an exhaustive effort to find optimal conditions and methods for a high throughput amenable screen to identify small molecule inhibitors of the methylated DNA reader protein MBD2 (*110*). We now apply our optimal assay to the National Institutes of Health (NIH) Molecular Libraries Probe Production Centers Network (MLPCN) small molecule screen of ~370,000 compounds. Four small molecules are identified as effectively disrupting the MBD2-meDNA interaction with micromolar affinity. We go on to show at least 2 of these compounds have biological activity, causing gene re-expression of known DNA methylation silenced genes.

4.3 Results

Primary Screen To Find MBD2(MBD)-meDNA Interaction Inhibitors

The MBD2-methylated DNA TR FRET assay described in chapter 3 was provided to The Scripps Research Institute® in Florida. With some optimization they further miniaturized the assay from 384 to 1536 well plate format. The assay was then subjected to the entire MLPCN screen comprising 370,276 compounds, as described in the methods. The screen deemed 1,149 compounds (0.31%) as significant MBD2-meDNA interaction inhibitors: with significance defined as loss of FRET signal >3 standard deviations from the median (dotted red line Figure 4.1 A and B). The Z' factor for the primary screen was calculated at 0.84±0.01 and Z=0.79±0.10.

Only 1,036 of the compounds from the primary screen were available for further work. Each was screened at a single dose in triplicate to identify false positives from the primary screen. Only 271 of the 1,036 compounds (26.2%) in this second screen validated as significant inhibitors of the MBD2-meDNA interaction (Figure 4.2A). The Z' factor for the secondary screening was 0.89±0.03 (Z=0.27±0.00). We expected the Z factor statistic in the secondary screen to be low, as all of the compounds had already been vetted in the primary effort and were predicted to be significant inhibitors of our assay. In an attempt to use

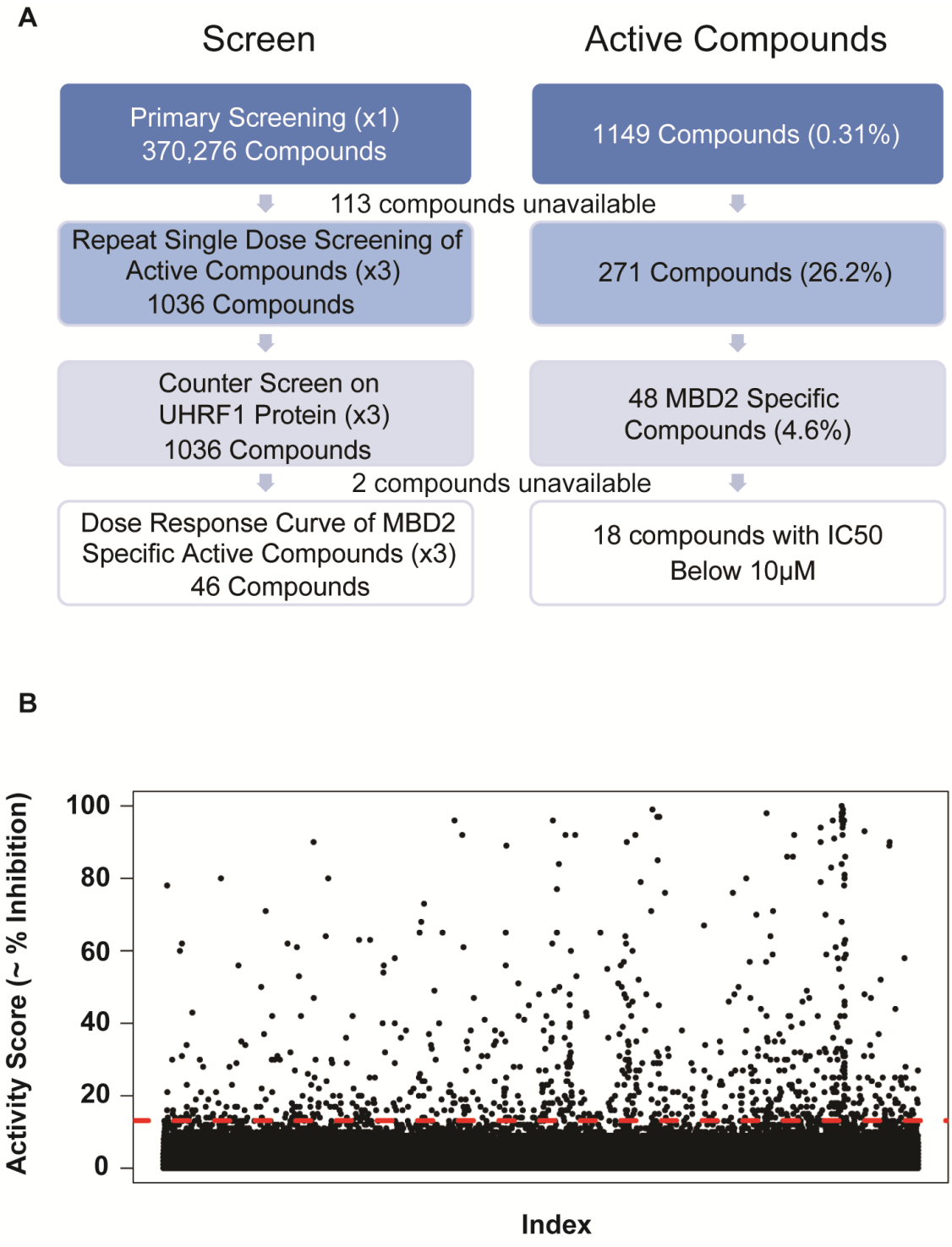


Figure 4.1: Primary screen results for the MBD2-meDNA TR FRET high throughput screen.

(A) Summary of all experiments and resulting active compounds for each step of the MLPCN screen. **(B)** Results of the single dose single well primary screen for all 370,276 compounds. Significance cutoff is indicated by the doted red line.

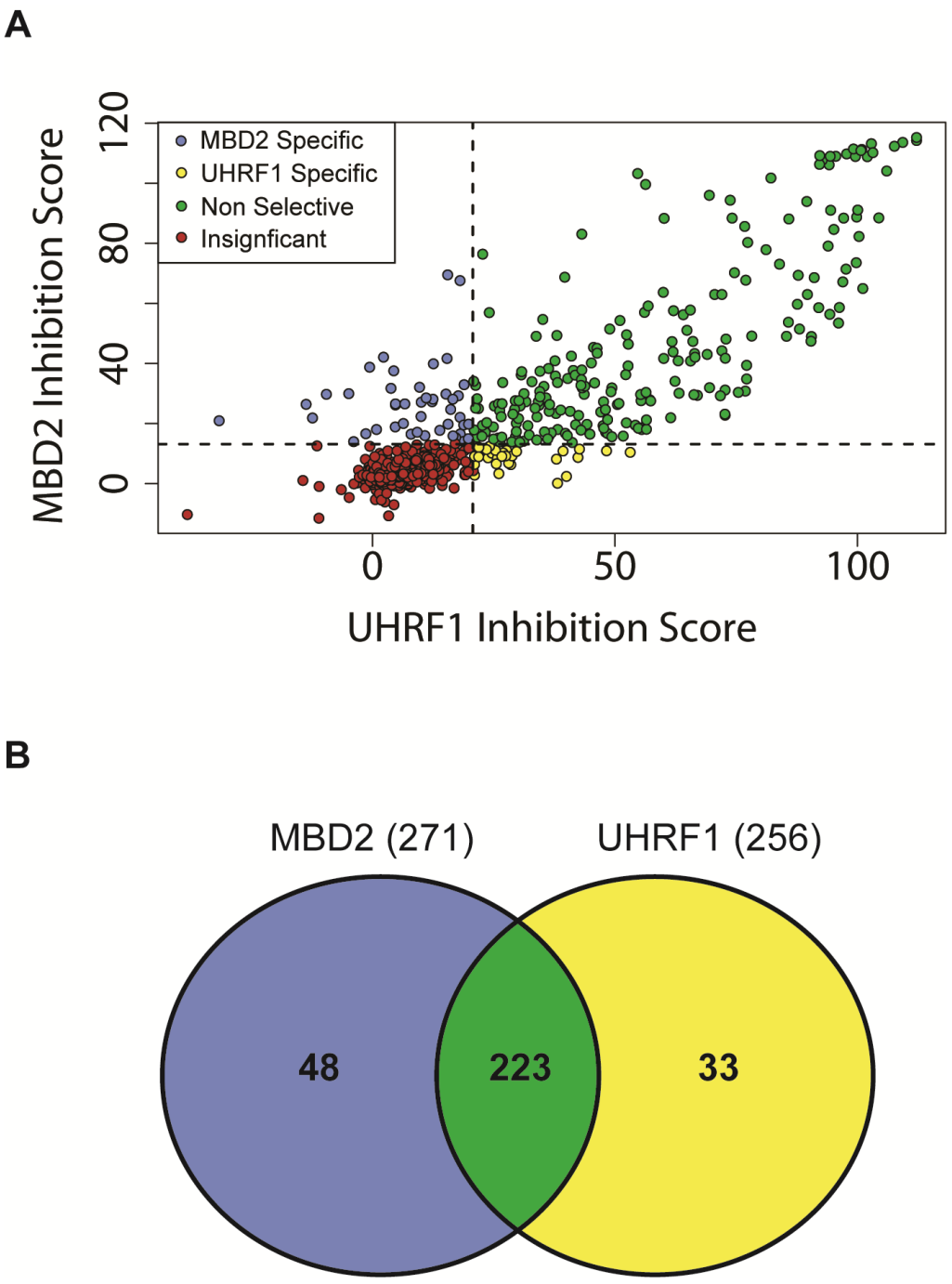


Figure 4.2: Summary of secondary screening results.

(A) Results of the secondary screening, single dose in triplicate, on the 1036 active compounds identified in the primary screen. Compounds are plotted relative to their inhibition of MBD2 (y axis) and UHRF1 (x-axis). Compounds which only significantly inhibit MBD2 are labeled in blue. **(B)** Venn diagram of secondary screen results.

our counter screen data as both a query for nonspecific inhibitors of MBD2-MBD and also interesting small molecules themselves we opted to use the SRA (methylated DNA reading) domain of UHRF1 protein in place of SP-1, the transcription factor used during assay development. We applied the 1,036 compounds from the MBD2-MBD primary screening effort in triplicate to a TR-FRET based assay using UHRF1-SRA and hemi-methylated DNA. The results showed that only 48 of the 271 small molecules that significantly inhibited MBD2-MBD after the triplicate screening were selective (Figure 4.2 A and B). The UHRF1-SRA TR FRET screen performed well with a Z'=0.79±0.10. The Z factor was 0.06±0.13, providing a good statistical confirmation of what is obvious in the results: that the majority of inhibitors from the MBD2-MBD screening are non-specific protein-DNA interaction disruptors.

Dose Response Series on Potential MBD2-MBD Inhibitors

Of the 48 MBD2-meDNA specific active compounds identified by the secondary screen we obtained 46 for further work. We then subjected each to the MBD2 and UHRF1 TR FRET assay in a dose response in triplicate. This experiment identified 18 compounds as both MBD2 specific and also having an IC_{50} below 10μM (Table 5.1). A more detailed look at the data suggests that many of the compounds have little

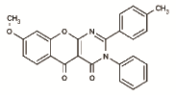
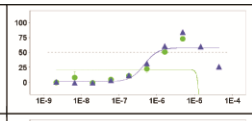
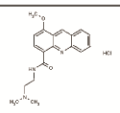
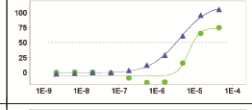
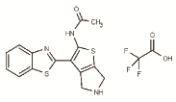
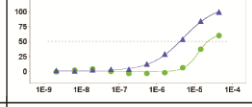
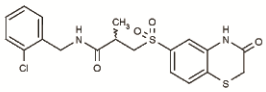
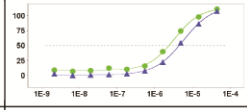
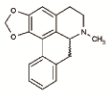
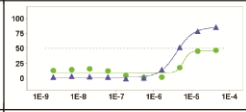
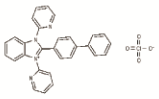
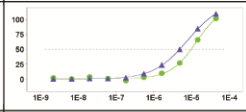
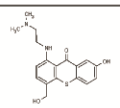
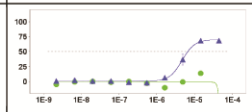
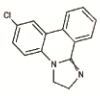
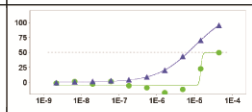
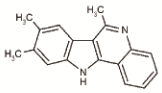
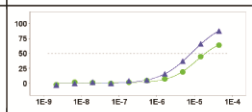
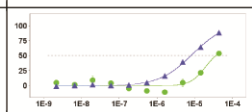
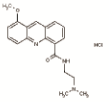
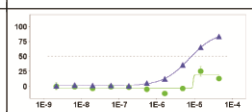
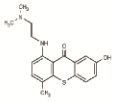
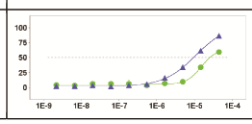
PubChem CID	JHH Identifier	Structure		IC50 (M)		Promiscuity	
				MBD2-MBD	UHRF1-SRA	Active	Assays Queried
1872967	KCC-107			988E-9	541E-9	5	537
386514	KCC-134			3.2E-6	9.1E-6	81	352
60138096	KCC-120			4.1E-6	22.4E-6	5	15
20923253	KCC-131			4.4E-6	2.4E-6	31	489
235224	KCC-123			4.7E-6	43.8E-6	68	535
16196160	KCC-116			4.8E-6	9.6E-6	18	513
333459	KCC-102			6.6E-6	43.8E-6	61	344
285027	KCC-112			6.7E-6	43.8E-6	20	192
5412959	KCC-128			8.E-6	19.E-6	39	266
285026	KCC-117			8.1E-6	37.E-6	27	258
386511	KCC-100			8.2E-6	43.8E-6	47	333
432873	KCC-104			9.4E-6	25.4E-6	46	268

Table 4.1: Summary of dose response on 17 active compounds identified in the primary and secondary screening.

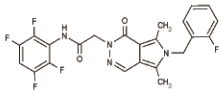
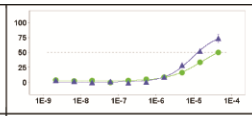
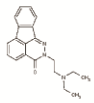
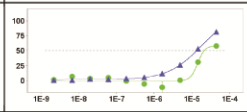
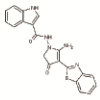
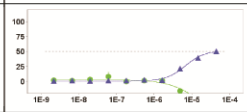
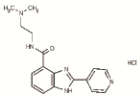
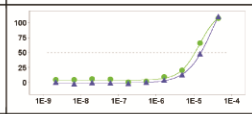
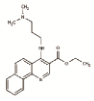
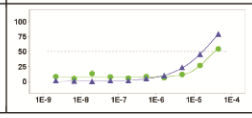
PubChem CID	JHH Identifier	Structure	Dose-response curves	IC50 (M)		Promiscuity	
				MBD2-MBD	UHRF1-SRA	Active	Assays Queried
16194171	KCC-114			12.6E-6	43.8E-6	49	636
2052803	KCC-111			13.4E-6	21.2E-6	70	645
2353771	KCC-127			14.6E-6	43.8E-6	19	617
386523	KCC-101			15.4E-6	10.6E-6	19	318
1561354	KCC-135			17.1E-6	35.7E-6	36	499

Table 4.1 Continued: Summary of dose response on 17 active compounds identified in the primary and secondary screening.

difference in their inhibition of UHRF1-SRA and MBD2-MBD, with only 5 having a ratio of their UHRF1 IC_{50} to MBD2 IC_{50} > 5. To this point we had only queried the specificity of these compounds with one non MBD2 methylation reader protein, UHRF1-SRA. To further investigate the specificity with other proteins we expressed, purified and tested multiple MBD family members and verified the recombinantly made protein preferentially bound methylated DNA (Figure 4.3). We then applied our TR FRET assay using MBD1-MBD, MeCP2-MBD, MBD2-MBD, UHRF1-SRA and transcription factor SP-1 to the 17 available compounds validated in the high throughput dose response screen (Figure 4.4). The results placed compounds into one of three categories: MBD selective (Figure 4.4 top panel, 7 total), MBD2 and MeCP2 selective (Figure 4.4 middle panel, 3 total), or non-selective (Figure 4.4 bottom panel, 7 total). All active compounds had IC_{50}'s in the micromolar range, confirming previous dose response data obtained from the Scripps Institute (Table 4.1 and Figure 4.4).

Epigenetically Silenced Gene Re-expression Potential of MBD2 Active Compounds

To lower the likelihood of a false negative we included all 17 compounds identified in the HTS dose response in our initial

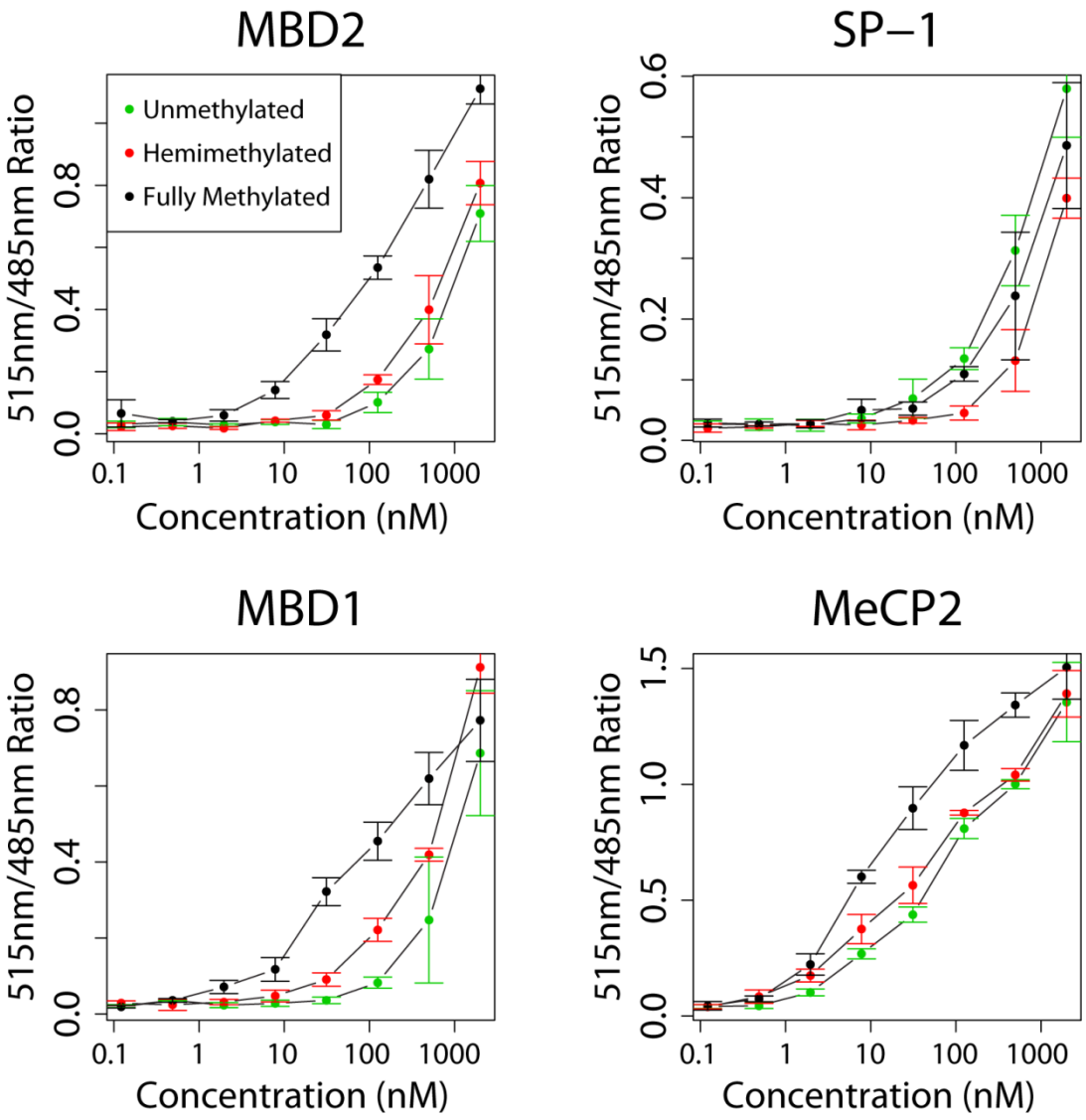


Figure 4.3: TR-FRET binding curves for MBD family members to DNA.

MBD family members were expressed in bacterial cells and recombinant protein was assayed against fully, hemi and unmethylated DNA using a TR-FRET assay.

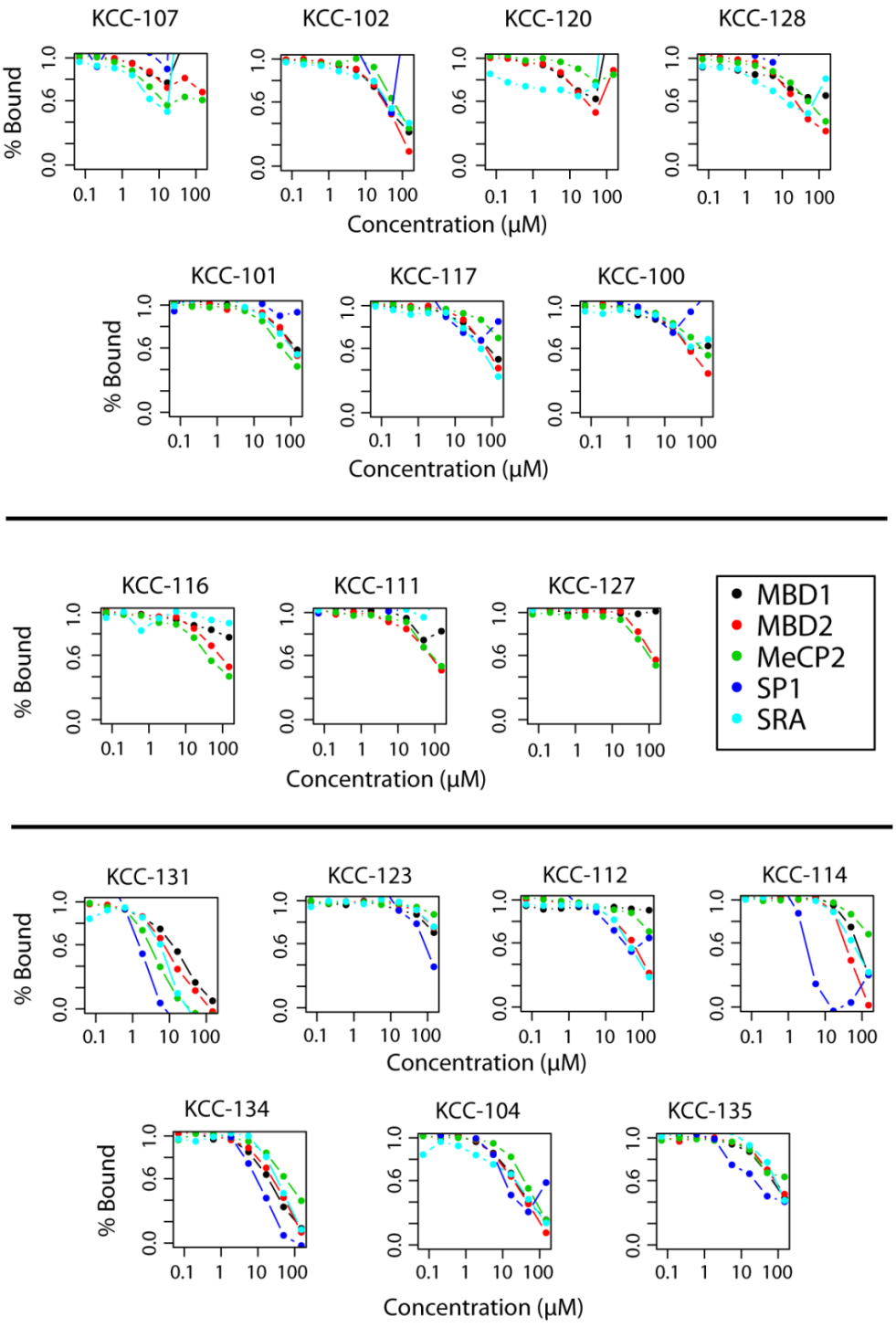


Figure 4.4: Dose response curves for MBD family members with HTS active compounds.

All 17 active small molecule inhibitors identified by the HTS screen were added to MBD family members in a dose response series to assay specificity. Results binned small molecules into one of three categories: MBD specific (upper group), MeCP2 and MBD2 specific (middle group) or non-specific (lower group).

examination of biological activity, regardless of their selectivity when tested against the other MBD family members. We began by evaluating the compounds' abilities to re-express known DNA methylation silenced genes: GSTP1 in MCF7 cells (breast cancer) as well as GSTP1 and PTGS2 in LNCaP cells (prostate cancer). Compounds were evaluated at a single dose for 48 hours (Figure 4.5A). A total of 6 compounds relieved epigenetic silencing in at least 2 of the 3 genes tested (black arrows Figure 4.5A).

To confirm these results HEK293 cells containing M.SssI treated GSTP1 promoter (fully methylated) driving a luciferase reporter gene were exposed to compound (Figure 4.5B, Table 4.2). An unmethylated CMV driven renilla served as a transfection and viability control. Encouragingly, 5 out of 6 of the compounds which re-expressed endogenously silenced genes also showed significant luciferase expression at a minimum of 2 concentrations (black arrows Figure 4.5B). KCC-107 was unable to cause gene re-expression in the reporter assay despite causing endogenous gene re-expression. KCC-134 and 112 also relieved epigenetic repression in both experiments but had very similar family members already represented (KCC-102 and KCC-117 respectively). As such these compounds were not tested further.

To validate these findings we selected the 6 compounds that relieved epigenetic repression in single doses for further work (Figure

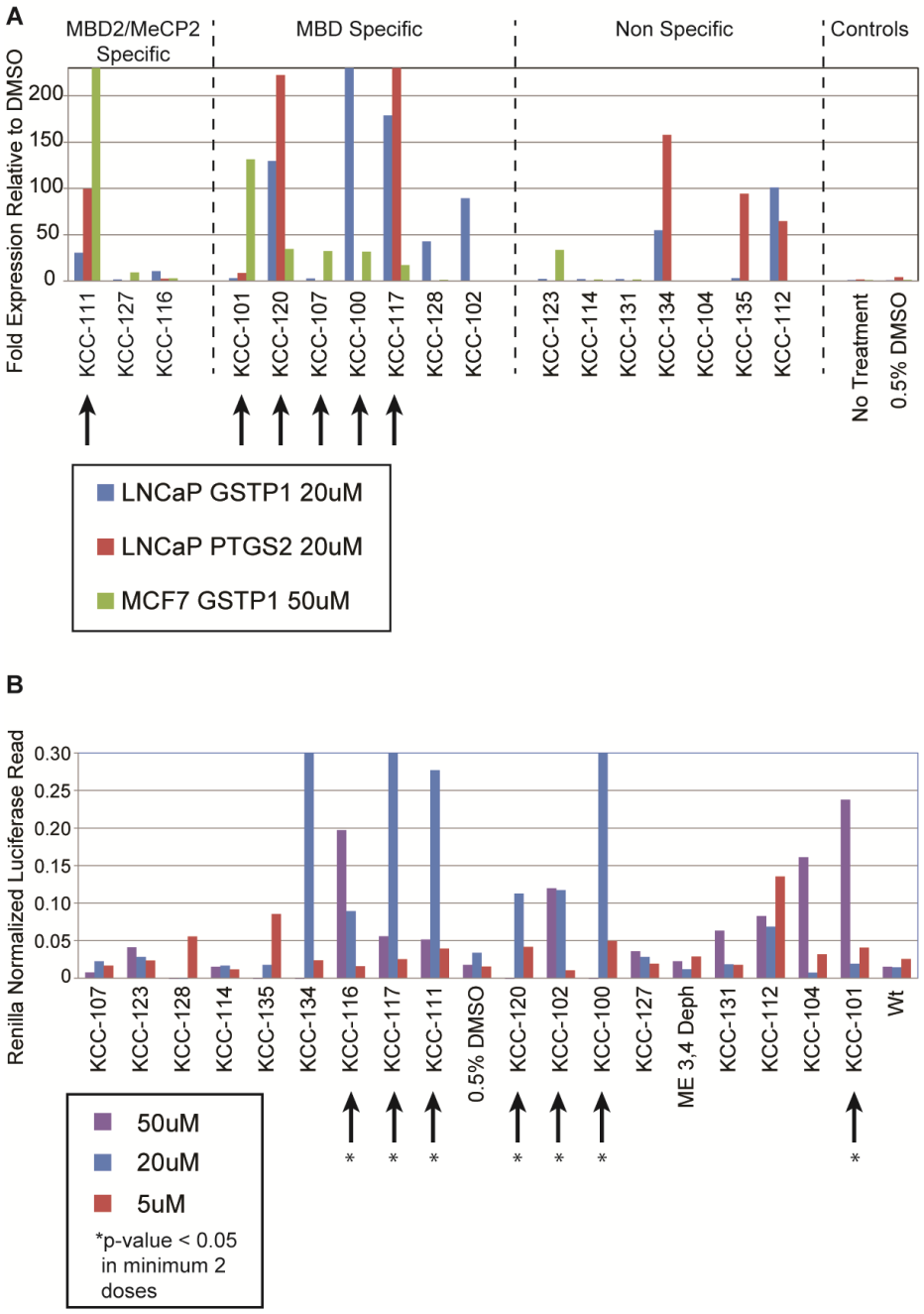


Figure 4.5: Gene re-expression potential of small molecules identified by high throughput screening.

(A) RNA transcript levels of endogenously methylated genes GSTP1 and PTGS2 measured by real time PCR following a 48 hour treatment of LNCaP or MCF7 cells at a single dose. Compounds indicated by the arrow had significant re-expression in at least 2 genes. **(B)** Firefly luciferase levels relative to a CMV-renilla control of HEK293 cells treated with a dose escalation series of small molecule inhibitors identified by the HTS. Luciferase expression was driven by a fully methylated GSTP1 promoter. Arrows indicate compounds which relieve epigenetic repression of the luciferase promoter at a minimum of two concentrations.

Sample	Dose	CMV-RFP Positive % Gated	CMV-RFP Negative % Gated	Combined % Gated
KCC-111	25μM	0.26	0	0.26
	12.5μM	0.53	1.95	2.48
	6.25μM	0.11	1.1	1.21
KCC-127	25μM	0.54	0.08	0.62
	12.5μM	0.42	0.25	0.67
	6.25μM	0.82	0.25	1.07
KCC-116	25μM	0.07	0.52	0.59
	12.5μM	0.18	0.37	0.55
	6.25μM	-	-	N/A
KCC-101	25μM	0	0	0
	12.5μM	0.21	0	0.21
	6.25μM	0.3	0.74	1.04
KCC-120	25μM	0.18	31.83	32.01
	12.5μM	0	16.73	16.73
	6.25μM	0	0.38	0.38
KCC-107	25μM	0.51	0.15	0.66
	12.5μM	0.69	0.08	0.77
	6.25μM	0.45	0.1	0.55
KCC-100	25μM	0.2	7.4	7.6
	12.5μM	0.28	2.25	2.53
	6.25μM	0	1.54	1.54
KCC-117	25μM	0.25	0.49	0.74
	12.5μM	0.38	2.3	2.68
	6.25μM	0.22	1.3	1.52
KCC-128	25μM	0	0	0
	12.5μM	0	0	0
	6.25μM	0	0.68	0.68
KCC-102	25μM	0	65.5	65.5
	12.5μM	0	20.86	20.86
	6.25μM	0	26.5	26.5
0.5% DMSO		0.99	0.05	1.04
No Treatment		0.66	0.01	0.67
No Transfection		0	0	0
both unmethylated		36.64	22.5	59.14

Table 4.2: Flow cytometry results of HTS active compounds with a methylated GSTP1 promoter driving GFP.

Percentage of cells with expression of a fully methylated GSTP1 promoter driving GFP The first column lists cells that also had expression of a CMV driven renilla while the second describes cells that were GFP (methylated promoter) positive but not RFP (unmethylated CMV promoter) positive. The final column is the combination of columns 1 and 2.

4.5), and also included 3 more small molecules which showed either MBD2/MeCP2 selectivity in the TR FRET MBD family screen (KCC-127 and KCC-116) or had family members that showed gene re-expression and also had slight activity themselves (KCC-102). We treated MCF7 cells for 48 hours in a dose response series and assessed expression levels of a number of genes previously established as epigenetically repressed, and sensitive to MBD2 knock down(*160*). KCC-120 and KCC-111 showed clear dose response re-expression of multiple epigenetically silenced genes, with expression levels 1-2 logs higher than DMSO treated cells (Figure 4.6).

Biophysical Interactions of Active Compounds with meDNA and MBD2 Protein

After establishing 2 of our 17 compounds as capable of relieving epigenetic repression in relevant cancer models we sought to confirm whether they were binding meDNA or MBD2-MBD protein. In an effort to prevent missing a possible protein binding small molecule we tested all 17 TR FRET based active compounds, not just those with biological activity, using a DNA intercalator competition assay. Briefly, either fully methylated or unmethylated plasmid DNA containing the GSTP1 promoter was pre-incubated with SYBR® green reagent and then incubated with increasing concentrations of small molecule for 1 hour.

Compounds which interact with DNA, either through minor groove or intercalation mechanisms, should disrupt SYBR® green binding to DNA and will result in a loss of fluorescence signal (see methods for details). Surprisingly, despite incorporating a counter screen to eliminate DNA binding molecules, the majority of active small molecules disrupted SYBR® green binding to DNA, suggesting they were interacting with DNA (Figure 4.7). Only 4 compounds, KCC-131, KCC-127, KCC-101 and KCC-107, did not interact with DNA. Both KCC-120 and KCC-111, the two biologically active small molecules from our *in vivo* gene expression studies, significantly disrupted SYBR® green binding to DNA. No difference was observed between the methylated and unmethylated plasmid results, suggesting the compounds were not sensitive to DNA methylation status.

To verify that the protein was not targeted by our small molecules we performed a protein thermal melt curve analysis. When interacting with small molecules many proteins resist melting, causing a shift in their calculated melt temperature(*103*). For a nanomolar or micromolar small molecule a T_m shift of approximately 5-10^{o}C or 2-5^{o}C, respectively, is expected (*156, 161*). None of the compounds found by our screen changed the thermal melting temperature of the MBD2-MBD protein (Table 4.3). As a control, and alternate path of investigation, we included NF449, a small molecule that was active in

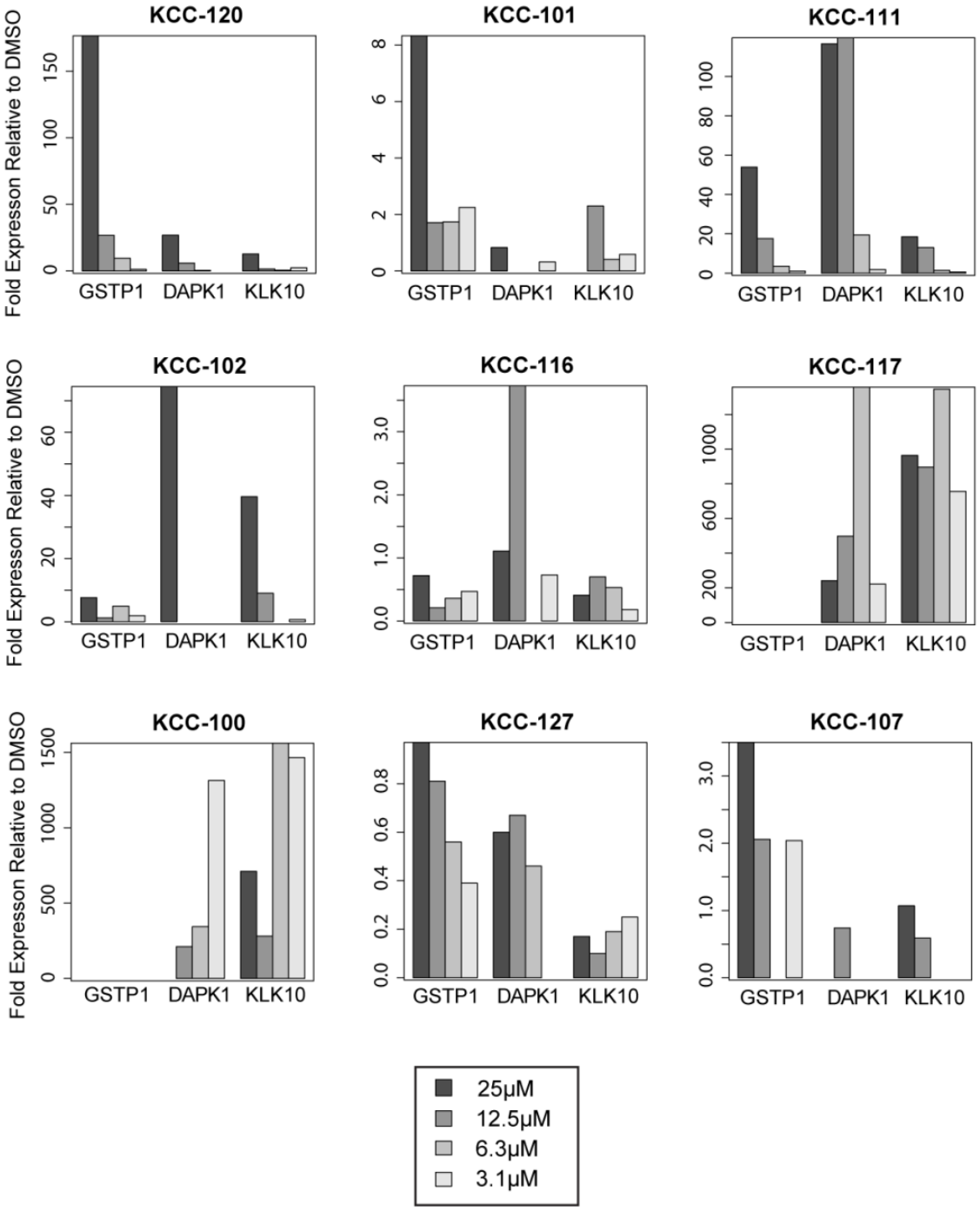


Figure 4.6: Dose response series re-expression of endogenously methylated genes in MCF7 cells

RNA expression levels of genes known to be sensitive to MBD2 siRNA knockdown after escalating dose treatment for 48 hours with small molecules identified by single concentration dosing. Note that y axis scale is different in all bar plots.

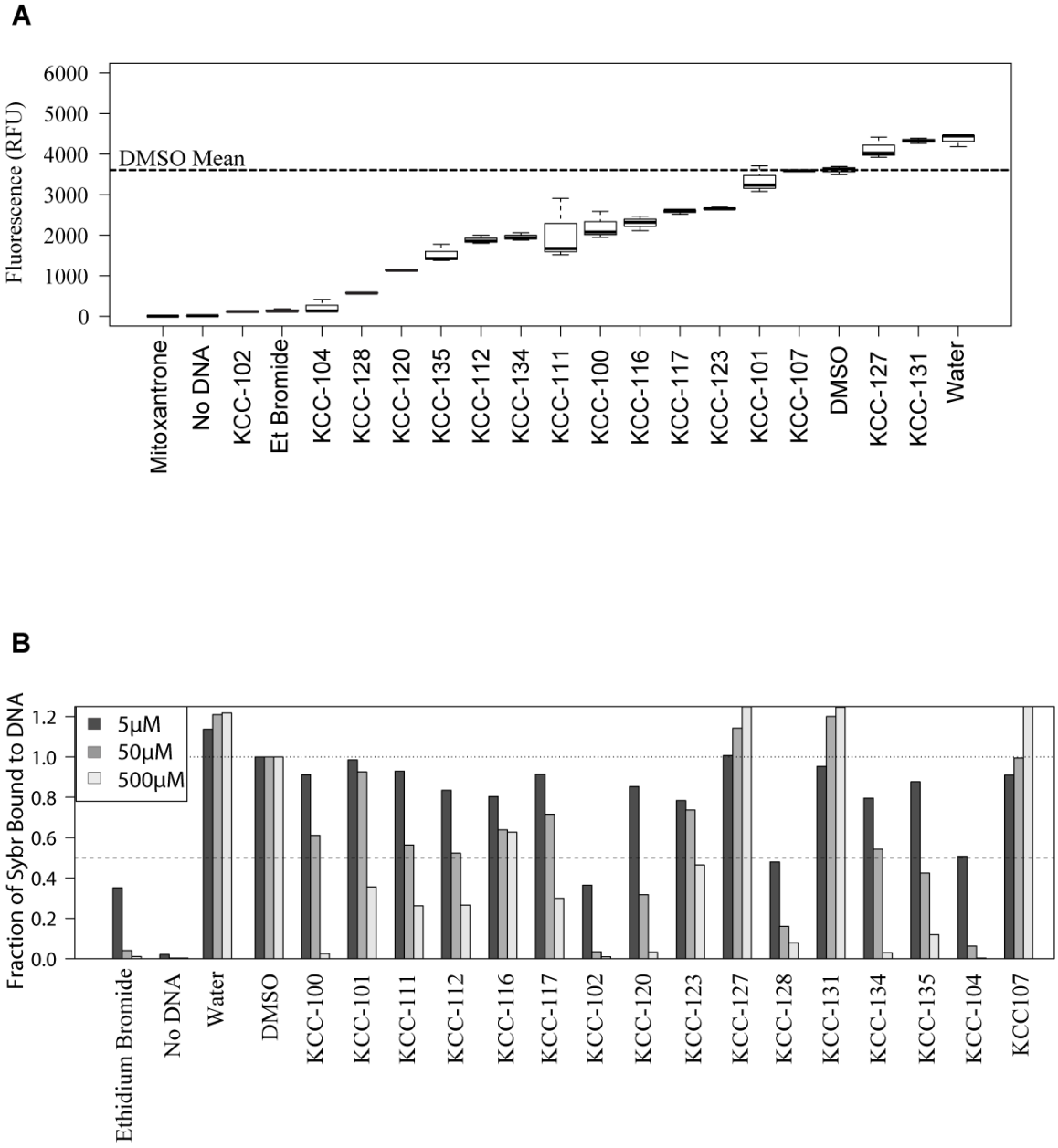


Figure 4.7: DNA intercalator completion assay on HTS active compounds

(A) Fluorescence measured with 50µM small molecule and 10x Sybr® green mixed with plasmid DNA. Dotted line represents DMSO control. Ethidium bromide (EtBr) served as a positive control. Compounds were run in triplicate and are sorted by median fluorescence intensity. **(B)** Fluorescence results of compounds at 5, 50 and 500µM with x10 Sybr® green.

	Melting Temperature (Celsius)			
Concentration (μM)	**50**	**25**	**12.5**	**6.25**
KCC-100	52.0	52.0	52.0	52.0
KCC-101	52.0	52.0	52.0	52.0
KCC-102	52.0	52.0	52.0	52.0
KCC-104	51.0	52.0	52.0	52.0
KCC-107	52.0	52.0	52.0	52.0
KCC-111	52.0	52.0	52.0	52.0
KCC-112	52.0	52.0	52.0	52.0
KCC-114	52.5	52.0	52.0	52.0
KCC-116	52.0	52.0	52.0	52.0
KCC-117	52.0	52.0	52.0	52.0
KCC-120	52.0	52.0	52.0	52.0
KCC-123	52.0	52.0	52.0	52.0
KCC-127	52.5	52.0	52.0	52.0
KCC-128	52.0	52.0	52.0	52.0
KCC-131	52.0	52.0	52.0	52.0
KCC-134	52.0	52.0	52.0	52.0
KCC-135	51.0	52.0	52.0	52.0
Suramin	53.0	52.5	52.0	52.0
NF023	53.0	52.0	52.0	52.0
NF449	52.0	52.0	52.0	52.0
NF110	52.0	52.0	52.0	52.0
0.5% DMSO	52.0	52.0	52.0	52.0
ME 3,4 Deph	52.0	52.0	52.0	52.0
Protein only	52.0	52.0	52.0	52.0

Table 4.3: Protein thermal melt temperatures with MBD2 and small molecule identified in the HTS and LOPAC™ screen.

our TR FRET assay pilot screen, and some of its family members (see chapter 3). Interestingly this family did cause a 1^{o}C shift in melting temperature, suggesting they could specifically bind MBD2.

Isothermal Calorimetry Analysis of MBD2-meDNA Active Compounds

To obtain more information about the biophysics of our novel MBD2-meDNA disrupting small molecules we performed isothermal calorimetry (ITC) measurements. We focused experiments on compounds KCC-120 and KCC-111 (able to relieve epigenetic repression *in vivo*), and KCC-127, KCC-131, KCC-101 and KCC-107 (not positive in our SYBR green displacement assay) and NF449 (positive hit from our pilot LOPAC$^{1280™}$ screen). KCC-102 was also included as a control. Both biologically active compounds, KCC-120 and 111, show strong binding to methylated DNA with K_d of ~40-50μM and 36μM respectively (Figure 4.8 and 4.9). This confirms the results of the DNA intercalator assay, implicating both of these compounds as DNA binding small molecules. KCC-101 and KCC-102 display significant heat release, comparable to that observed with KCC-111 and 120, when injected into methylated DNA (Figure 4.10 and 4.11). Though their curves did not follow a sigmoidal shape the background readings, compound into buffer, were <0.05μcal/sec on multiple replicates. Despite not observing saturable binding, the ITC data coupled with the DNA intercalator exclusion assay

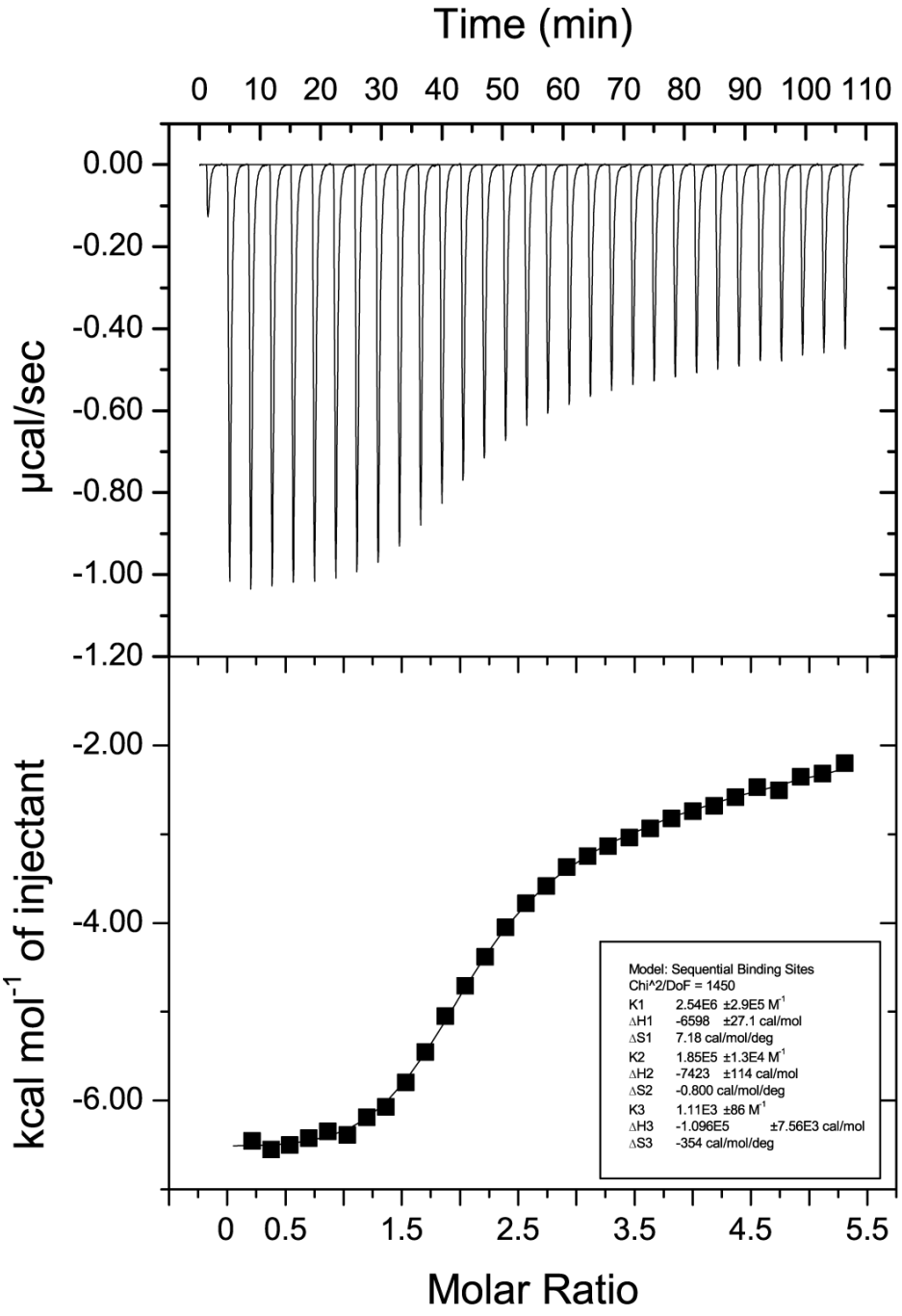


Figure 4.8: Isothermal calorimetry experiment with KCC-120

Small molecule injected into methylated DNA. The K_d is calculated at 40-50µM.

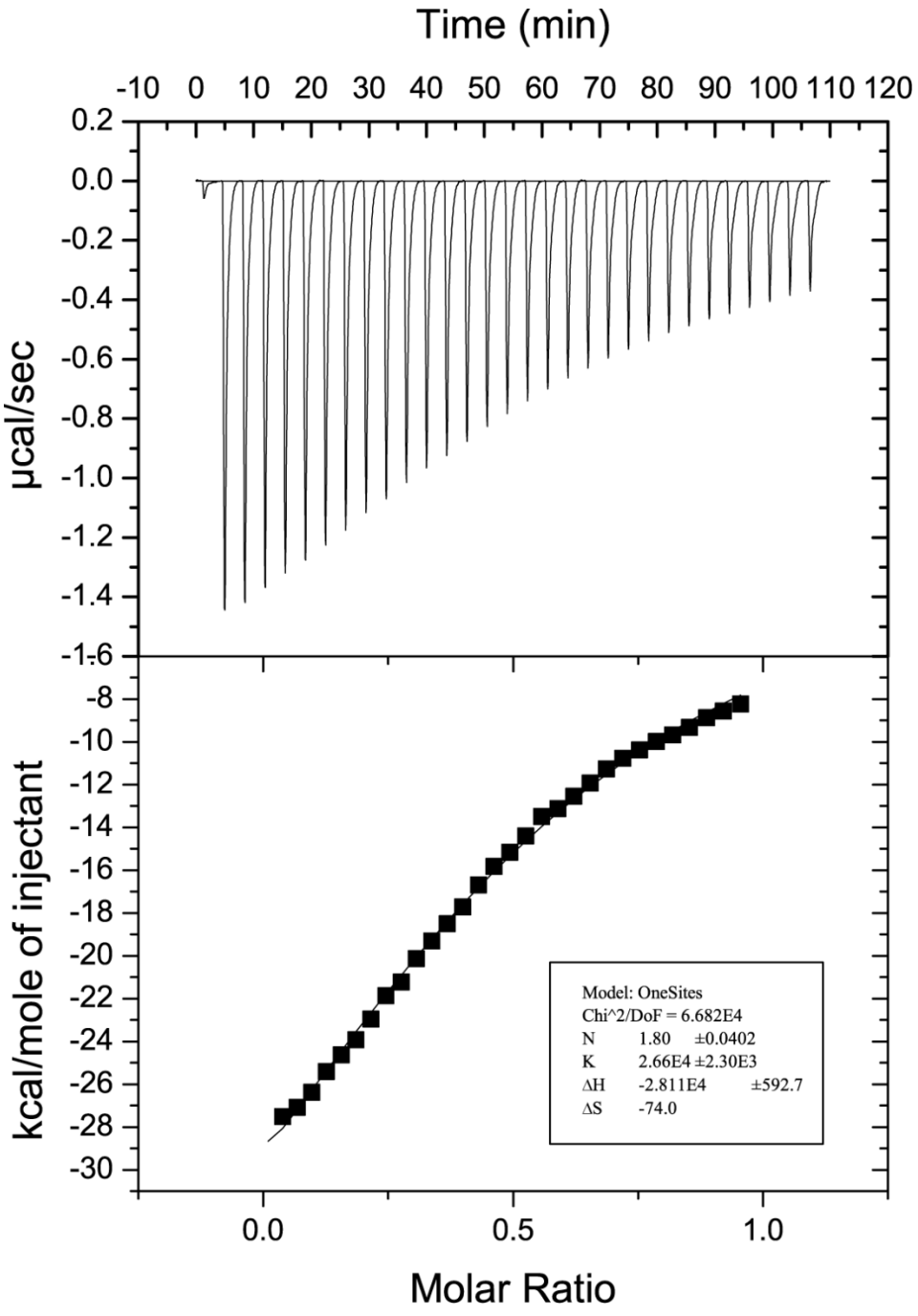


Figure 4.9: Isothermal calorimetry experiment with KCC-111

Small molecule injected into methylated DNA. The K_d is calculated at 36µM.

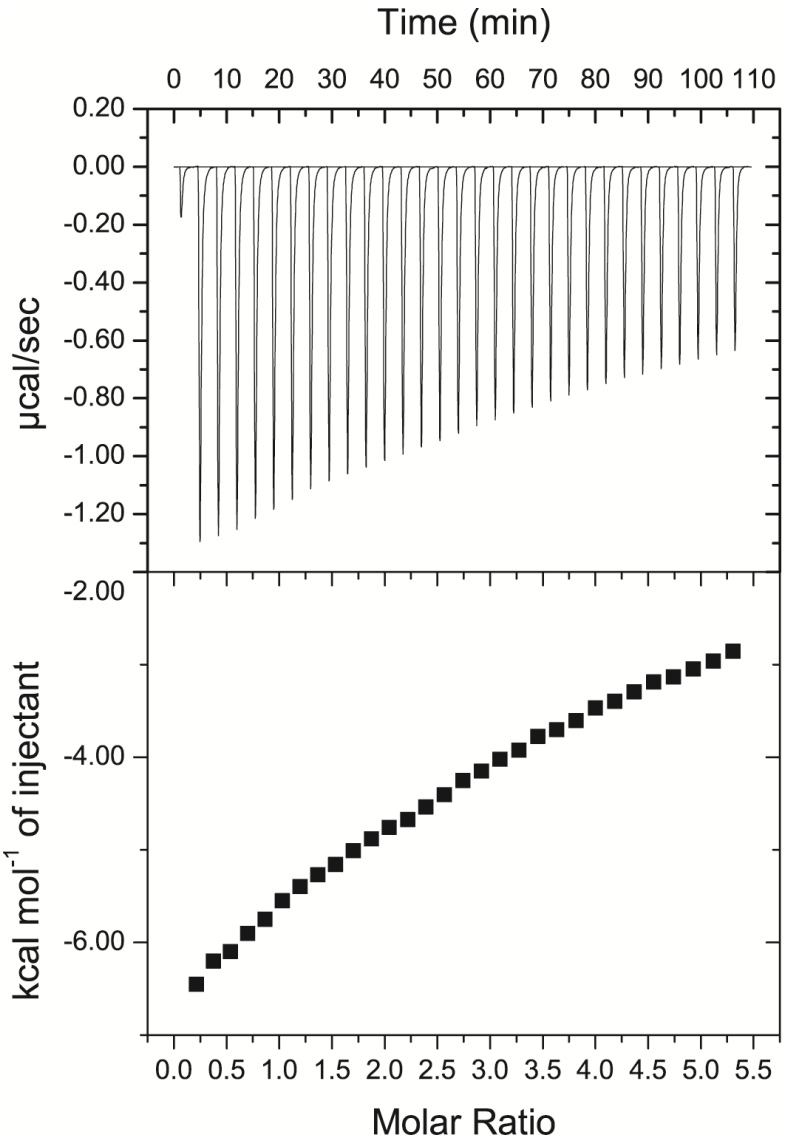


Figure 4.10: Isothermal calorimetry experiment with KCC-101

Small molecule injected into methylated DNA. Curve never saturates but significant heat release is present. Background (compound into buffer) showed negligible enthalpy results.

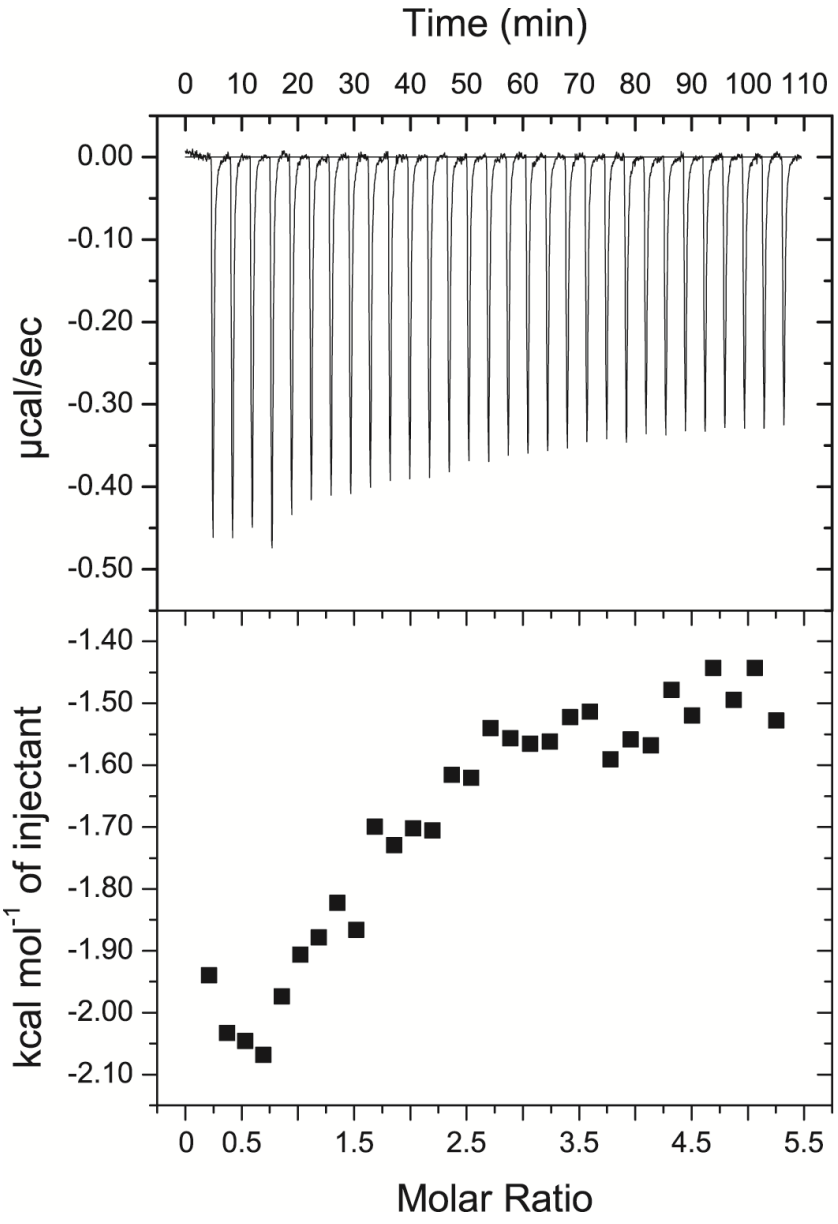


Figure 4.11: Isothermal calorimetry experiment with KCC-102

Small molecule injected into methylated DNA. Curve never saturates but significant heat release is present. Background (compound into buffer) showed negligible enthalpy results.

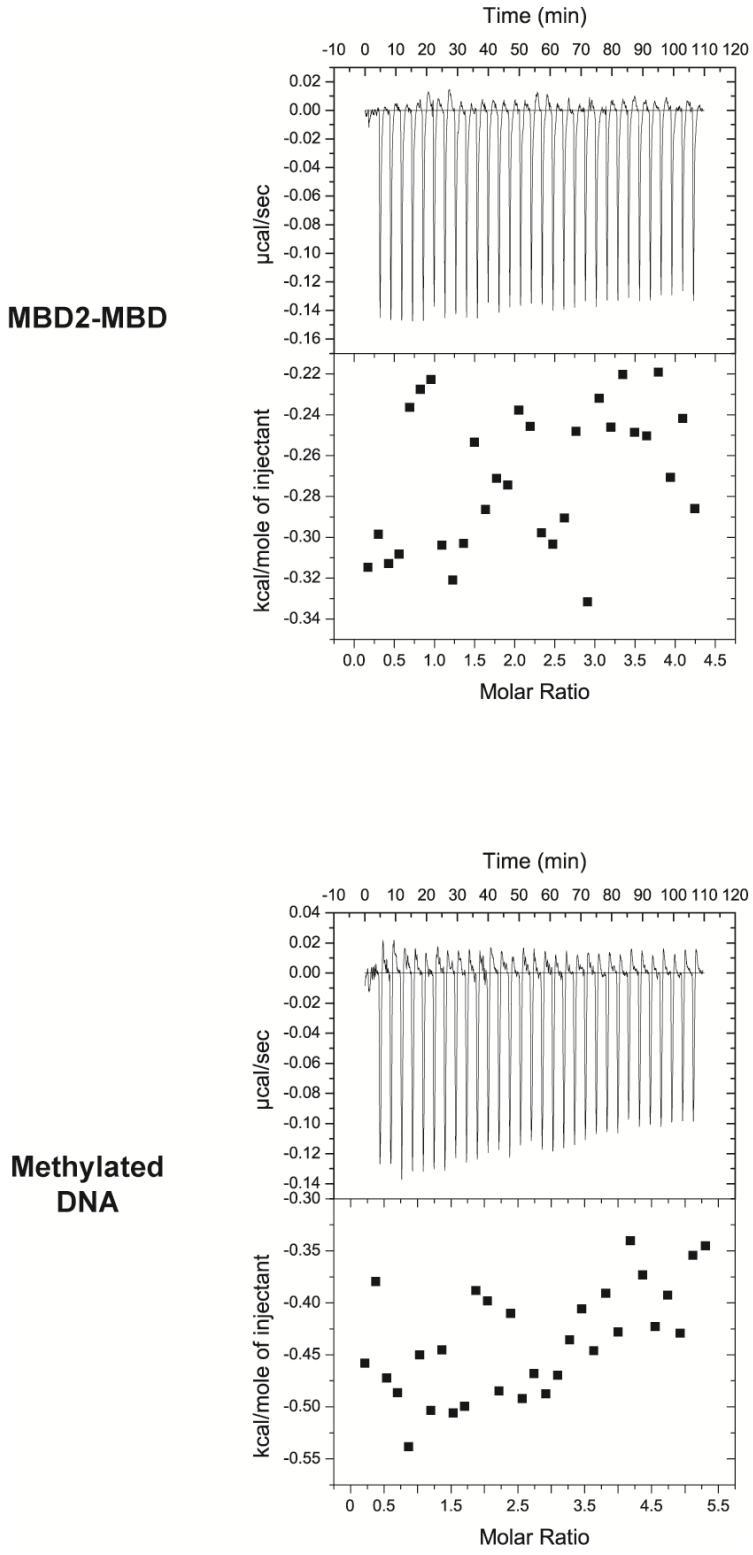


Figure 4.12: Isothermal calorimetry experiment with KCC-107

KCC-107 injected into MBD2 protein (upper panel) or methylated DNA (lower panel). Experiments were replicated at multiple temperatures and concentrations with similar results.

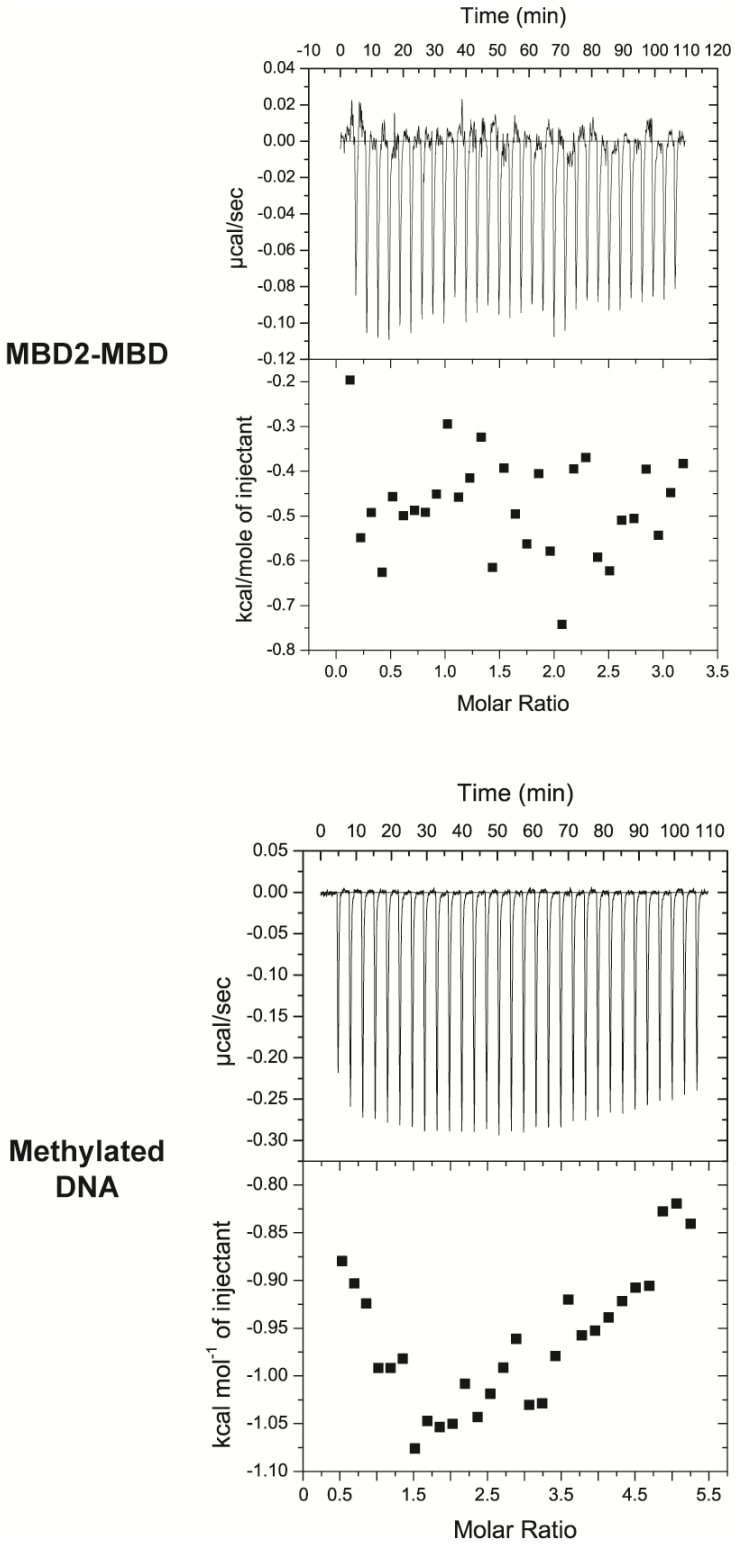
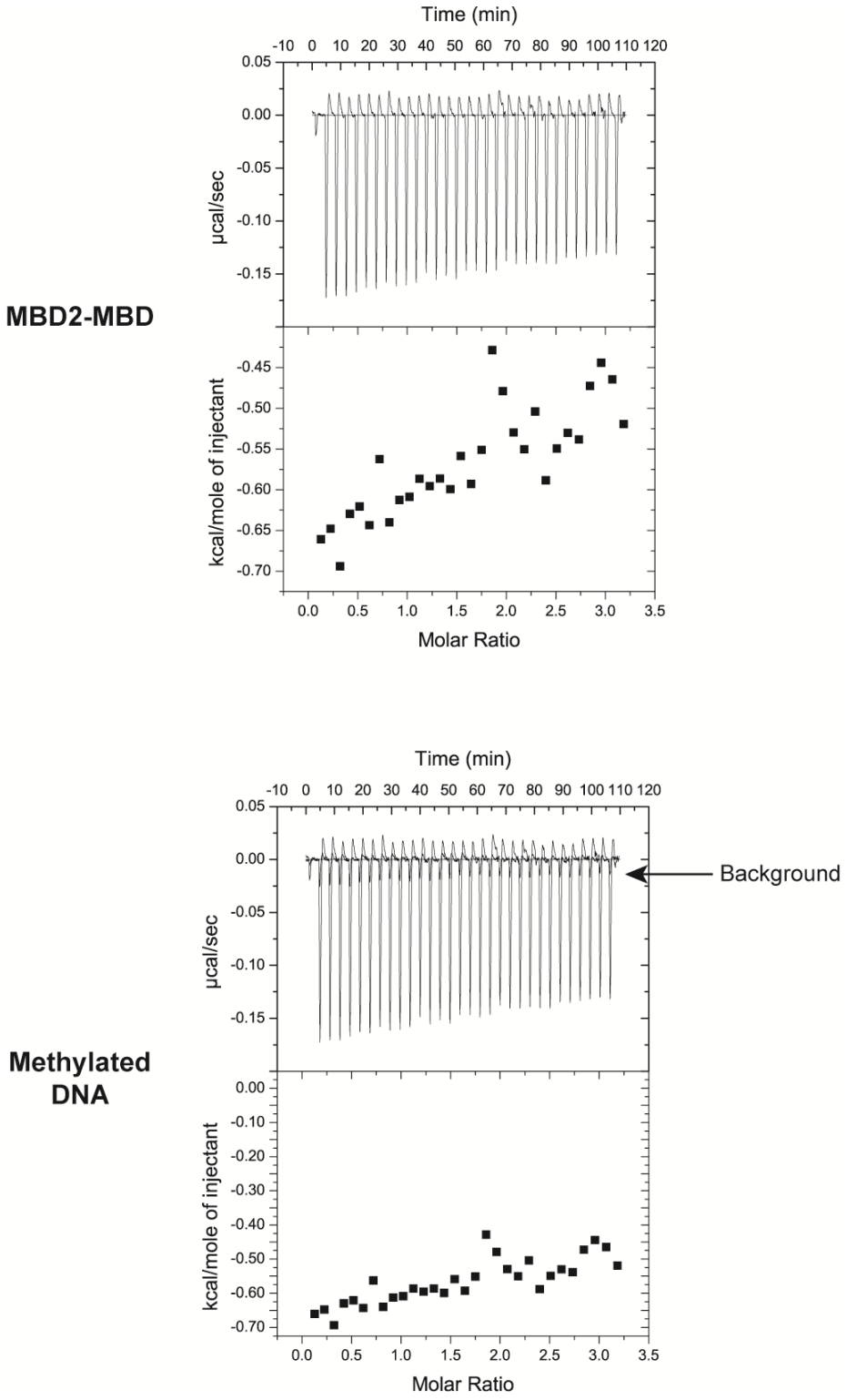


Figure 4.13: Isothermal calorimetry experiment with KCC-116

KCC-116 injected into MBD2 protein (upper panel) or methylated DNA (lower panel). Experiments were replicated at multiple temperatures and concentrations with similar results.



.

Figure 4.14: Isothermal calorimetry experiment with KCC-131

KCC-131 injected into MBD2 protein (upper panel) or methylated DNA (lower panel). Experiments were replicated at multiple temperatures and concentrations with similar results.

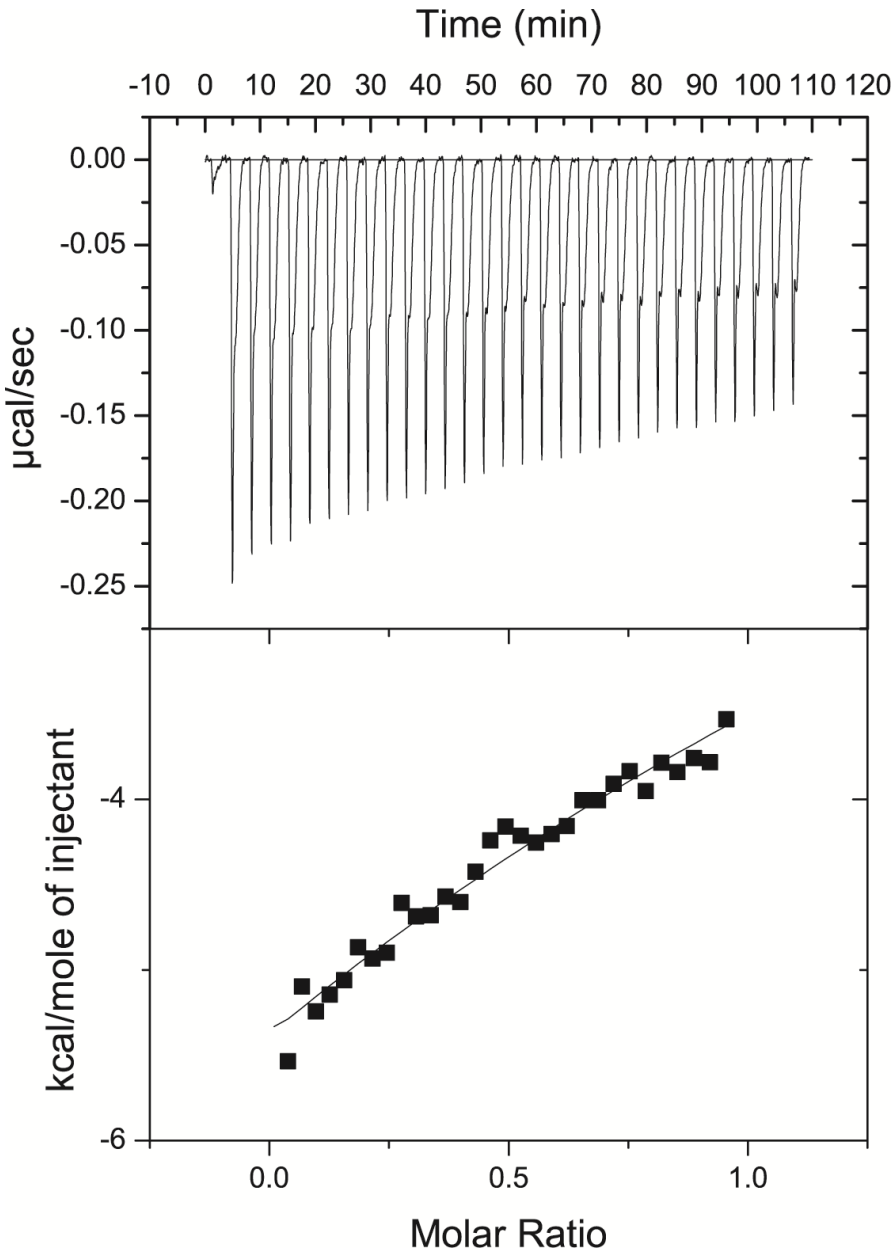


Figure 4.15: Isothermal calorimetry experiment with KCC-127

Methylated DNA injected into KCC-127. Experiments were replicated at multiple temperatures and concentrations with similar results. KCC-127 was insoluble at 50µM which may confound results presented here.

(Figure 4.7) make it highly likely that both of these compounds bind DNA.

KCC-107, KCC-116 and KCC-131 do not appear to bind either MBD2 protein or DNA in our ITC experiments (Figure 4.12-14). Similarly KCC-127 also does not appear to bind DNA (Figure 4.15). KCC-127 and 107 have strong optical effects at the wavelengths our TR FRET assay is read, which could explain why the assay identified them as active hits. KCC-116 and KCC-131, however, do not appear to have strong optical effects nor bind to either DNA or MBD2 protein in our ITC experiments. They may still interact with one of the two components in our assay but do so in such a weak manner, K_d > 40μM, that our thermal melt and ITC assays do not have the required sensitivity to measure them accurately. Interestingly, NF449 binds to MBD2-MBD protein in a dose dependent and saturable manner (Figure 4.16). The binding kinetics appear complicated, and suggest that there are two binding sites with K_d's of ~1.5μM and ~10μM, with the stronger interaction consistent with our pilot TR FRET results. Because of the complex nature of this binding, however, we had a difficult time fitting the results and thus we may be underestimating the K_d of this interaction. Control experiments of NF449 into methylated hairpin DNA or buffer demonstrate no appreciable binding or heat release, confirming this appears to be a protein specific interaction (Figure 4.17).

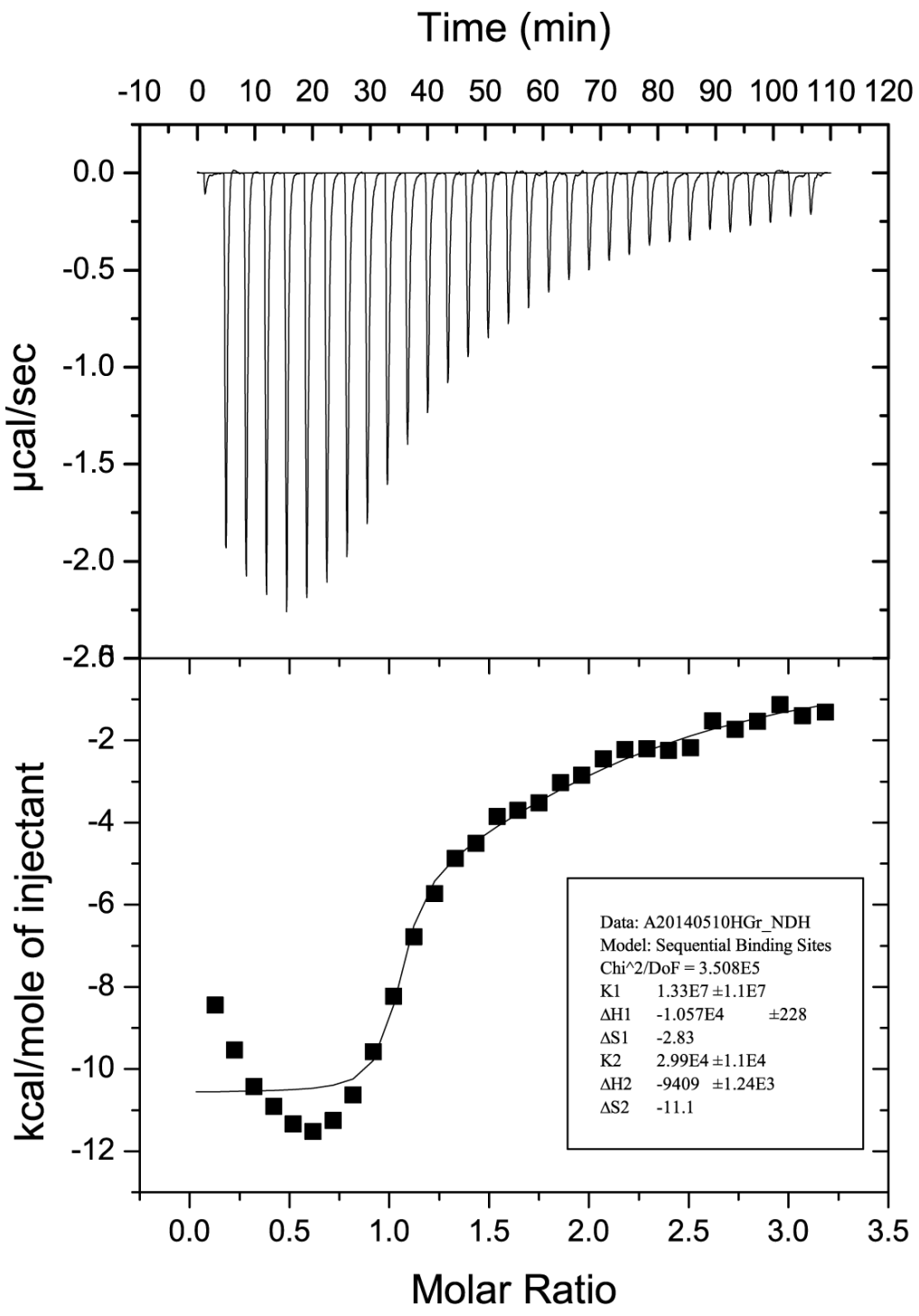


Figure 4.16: Isothermal calorimetry experiment with NF449 and MBD2-MBD

NF449 injected into MBD2 protein. Experiment was replicated with similar results suggesting two binding events occurring. The K_d of each is calculated at ~1.5**µ**M and ~10µM with the lower of the two consistent with the TR FRET results.

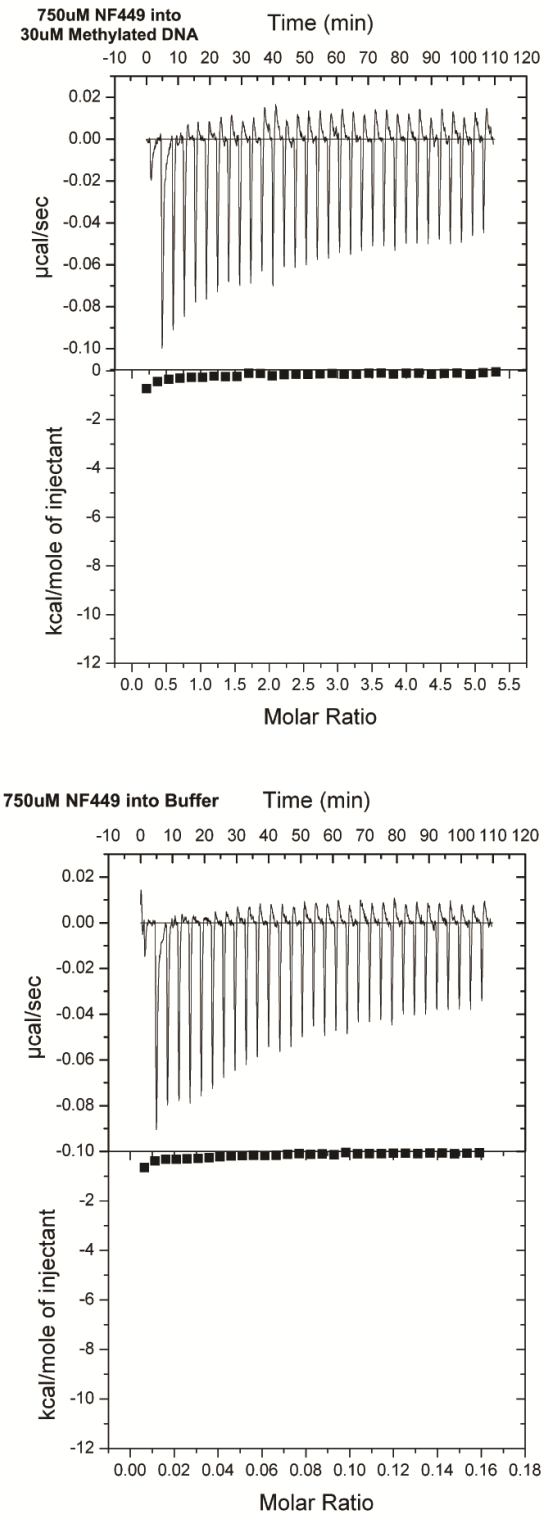


Figure 4.17: Isothermal calorimetry control experiments with NF449

NF449 injected into methylated DNA (upper) or buffer (lower) panel. Experiments were replicated with similar results.

Growth Curves With KCC Compounds

To make an initial assessment of the toxicity of these compounds we evaluated MCF7 breast cancer growth curves in the presence or absence of inhibitor. All of the compounds, except KCC-116 and 135, shown to interact with DNA by our previous work were quite toxic, with IC_{50}'s below 2μM (Figure 4.18A). The compounds which did not interact with DNA also did not appear to be toxic (Figure 4.18B). KCC-131 minimally interfered with cell growth at concentrations up to 50μM, with MCF7 growth areas at approximately 64.4% of control (Figure 4.19). Further work on KCC-107 and 116 suggested that though they did not appear to inhibit cell growth in this experiment, they did have significant toxicity at 50μM with 18.5% and 0.3% growth area relative to control respectively (Figure 4.19).

Chromatin Immunoprecipitation (ChIP) of Histone Modifications at Epigenetically Silenced Regions

With evidence that our two biologically active small molecules, KCC-120 and KCC-111, bind DNA we sought to show that the mechanism of this effect was by relief of epigenetic repression and not a simpler reason, such as steric hindrance. ChIP was performed with antibodies targeting a number of histone modifications as well as MBD2 and RNA polymerase II. We queried three sets of genes:

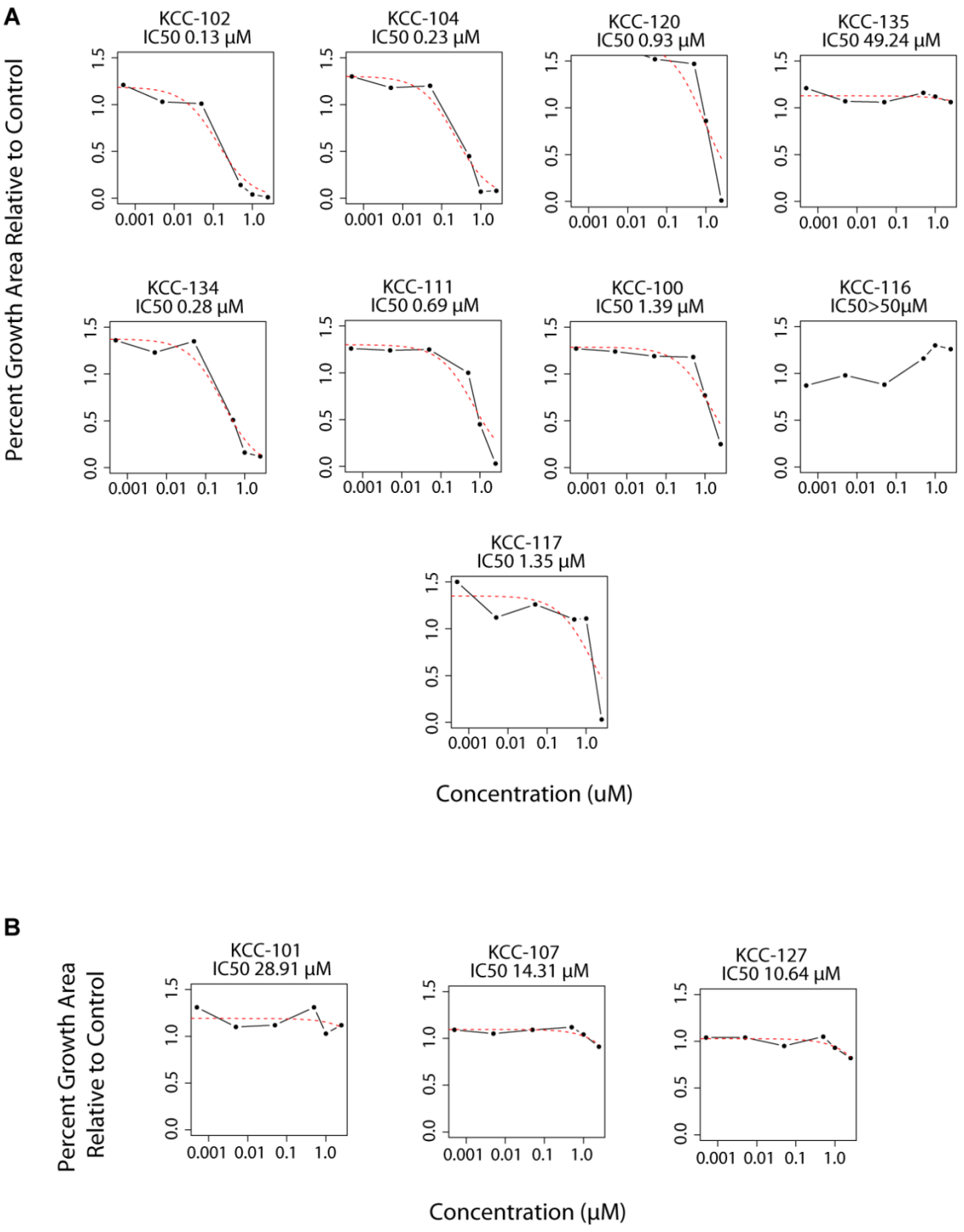


Figure 4.18: MCF7 growth curves in the presence of HTS assay active compounds

(A) Growth curves with compounds that interact with DNA based on the DNA intercalator experiment. All appear to be toxic at low micromolar concentration. **(B)** Growth curves with compounds that were inactive in our ITC and DNA intercalator competitions assay.

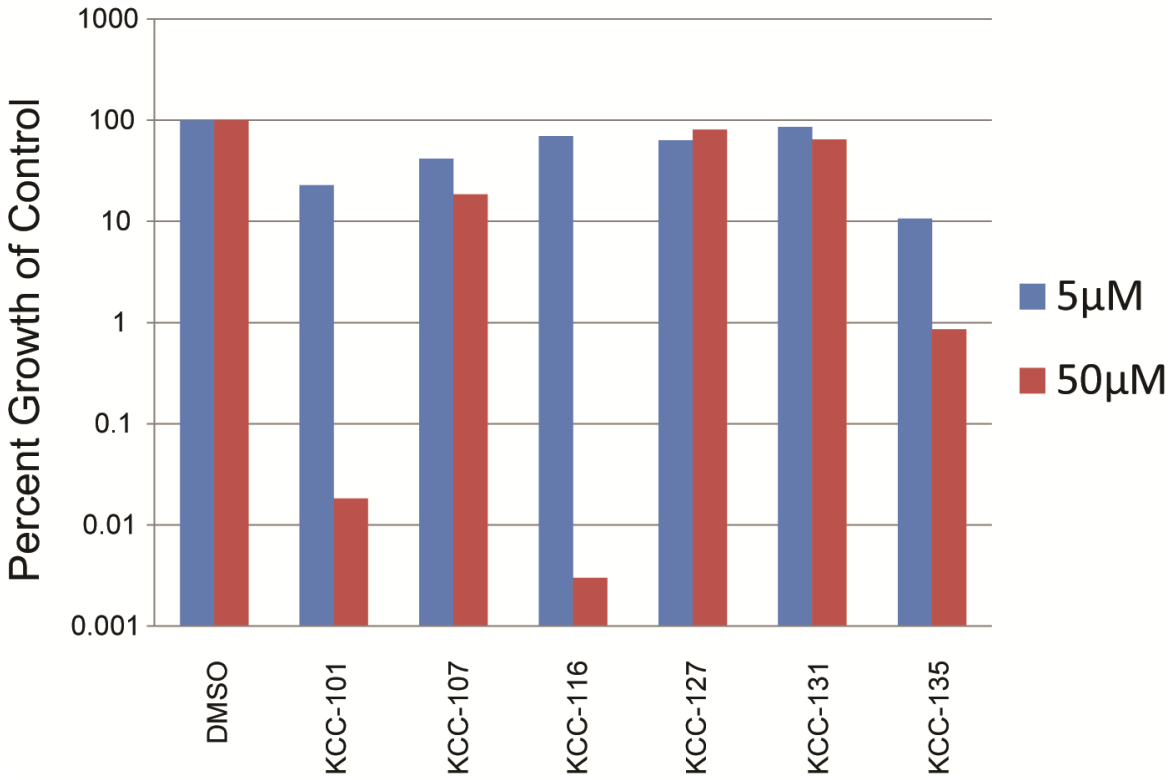


Figure 4.19: MCF7 growth curves in the presence of high concentration inhibitor

(1) Epigenetically silenced with a methylated promoter and low expression.

(2) Unmethylated promoter with low expression.

(3) Unmethylated promoter with high expression.

Unfortunately, due to technical issues, the RNA polymerase II precipitation universally showed little enrichment and the results were not included in the analysis.

Interestingly, genes with methylated promoters and low RNA expression showed histone changes consistent with a move towards a more open chromatin state (Figures 4.20 and 4.21). This included increases in histone H3K4 trimethylation and H3K9 acetylation marks associated with losses of H3K9 and H3K27 trimethylation. Genes with semi methylated or unmethylated promoters and low RNA expression displayed histone changes inconsistent with a change in the overall chromatin status, or shifted slightly towards a closed chromatin state (Figure 4.22 and 4.23). Likewise, genes with high expression, GAPDH and ESR1, exhibited states consistent with euchromatin in control samples, extremely enriched for H3K4 trimethylation and H3K9 acetylation, with little change after treatment with either compound (Figure 4.24). Surprisingly, the literature describes the gene MTAP as having low expression, but the chromatin state in our analysis is more consistent with genes of high expression. This discordance should be investigated further.

A summary view of the changes for all genes indicates that KCC-111 and KCC-120 effectively alter the histone code at methylated promoters, creating a more open chromatin state, while causing little effect at non methylated but silent or highly expressed promoter regions (Table 4.4). MBD2 ChIP, also shown here, may hint at a possible explanation for the mechanism behind this observation. In treated cells MBD2 levels appear lower at promoter regions, implying the compounds may selectively interfere with epigenetic machinery while allowing transcription factors and polymerases to access the site. These results, particularly those with the MBD2 protein, need to be repeated to confirm this trend.

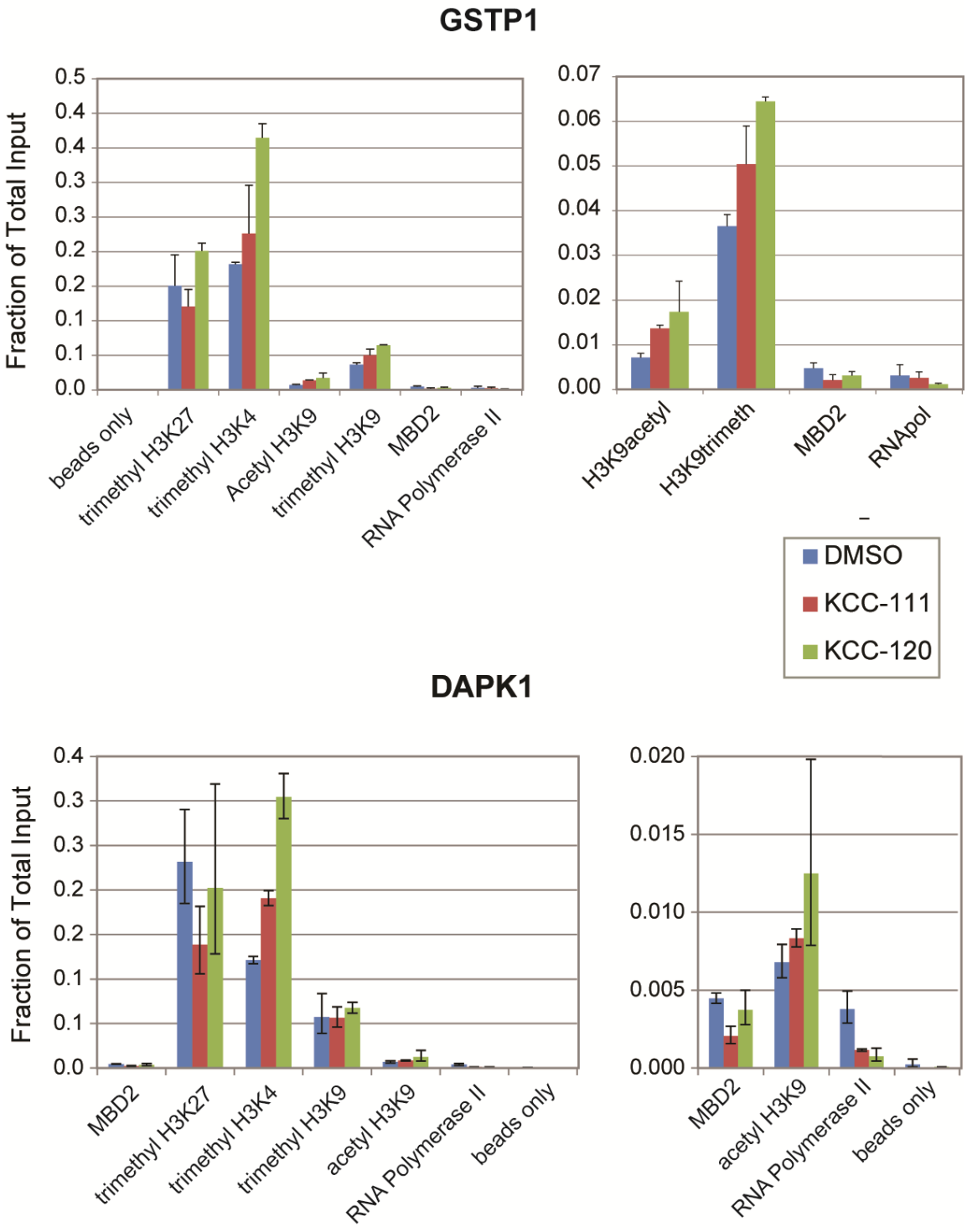


Figure 4.20: ChIP results of epigenetically controlled genes treated with KCC-120 and 111

Fraction of input for genes with methylated promoters and low RNA expression. RNA polymerase and MBD2 precipitations had universally low results. Bar plots on left and right for each gene are the same data plotted with different y axis scales.

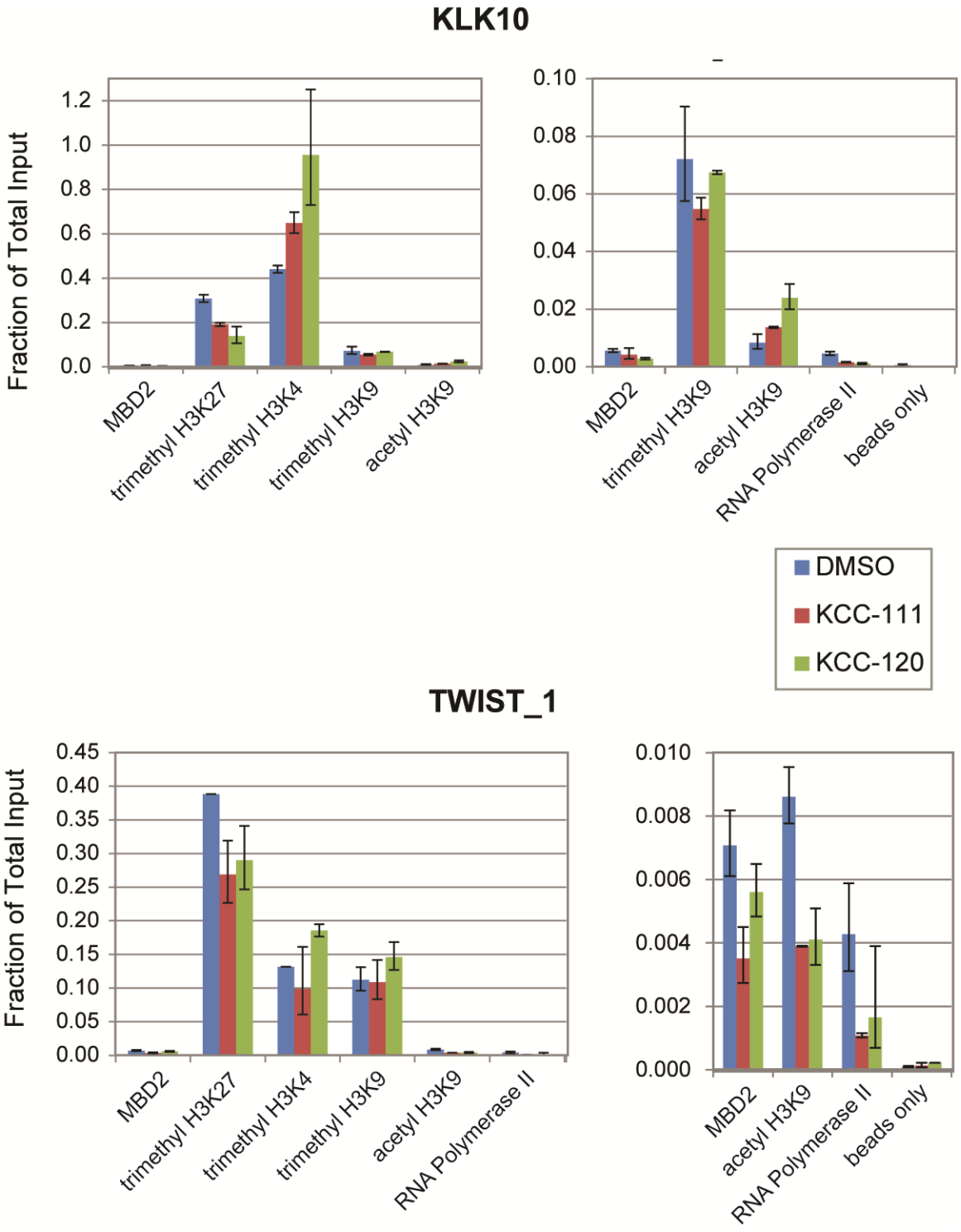


Figure 4.21 ChIP results of epigenetically controlled promoters treated with KCC-120 and 111 cont...

Fraction of input for genes with methylated promoters and low RNA expression. RNA polymerase and MBD2 precipitations had universally low results. Bar plots on left and right for each gene are the same data plotted with different y axis scales.

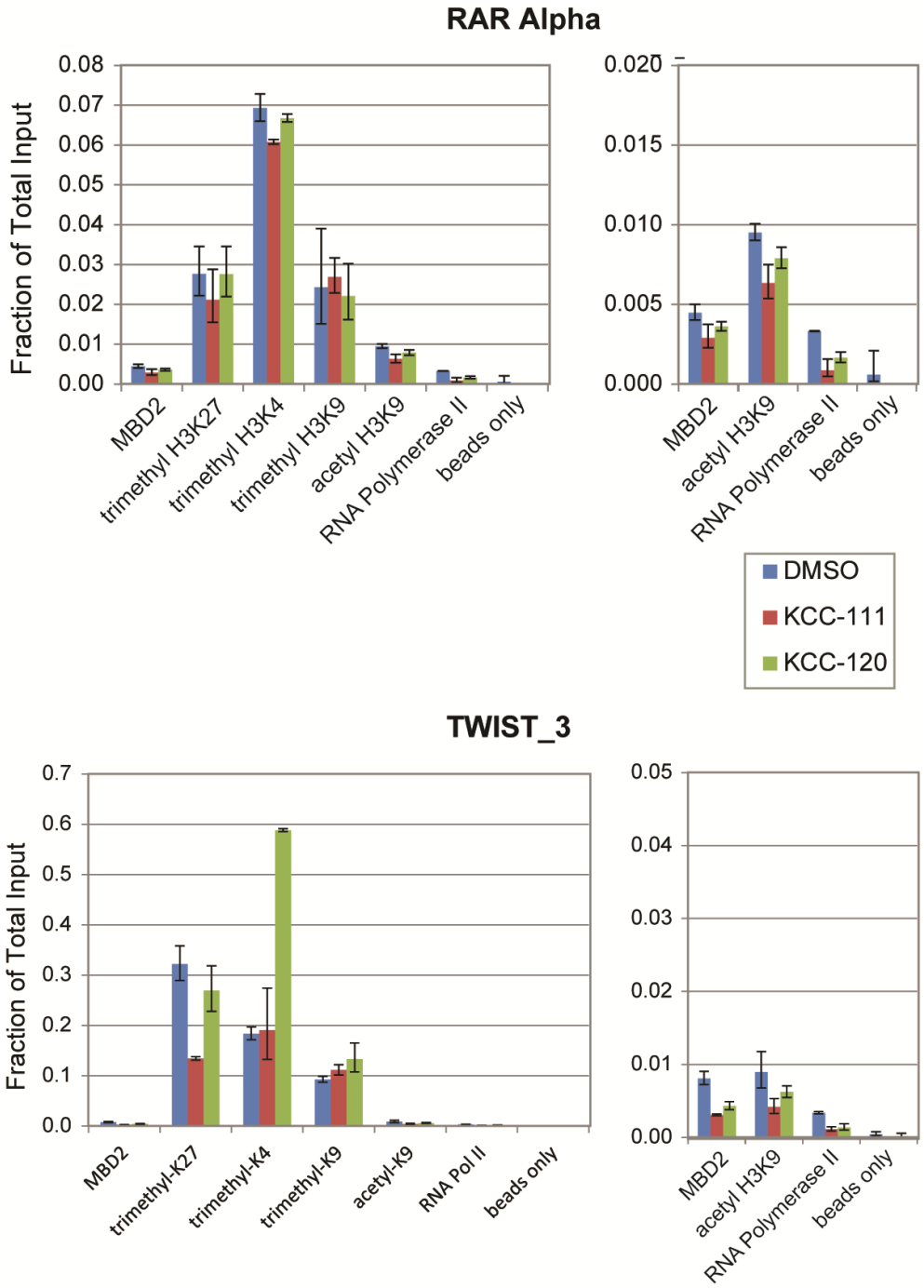


Figure 4.22: ChIP results of non-epigenetically controlled genes treated with KCC-120 and 111

Fraction of input for genes with semi (RAR Alpha) or unmethylated (TWIST_3) promoters and low RNA expression. RNA polymerase and MBD2 precipitations had universally low results. Bar plots on left and right for each gene are the same data plotted with different y axis scales. TWIST_3 primer set is 1000bp downstream of the first primer set (figure 4.21) and is in an unmethylated portion of the promoter region for TWIST.

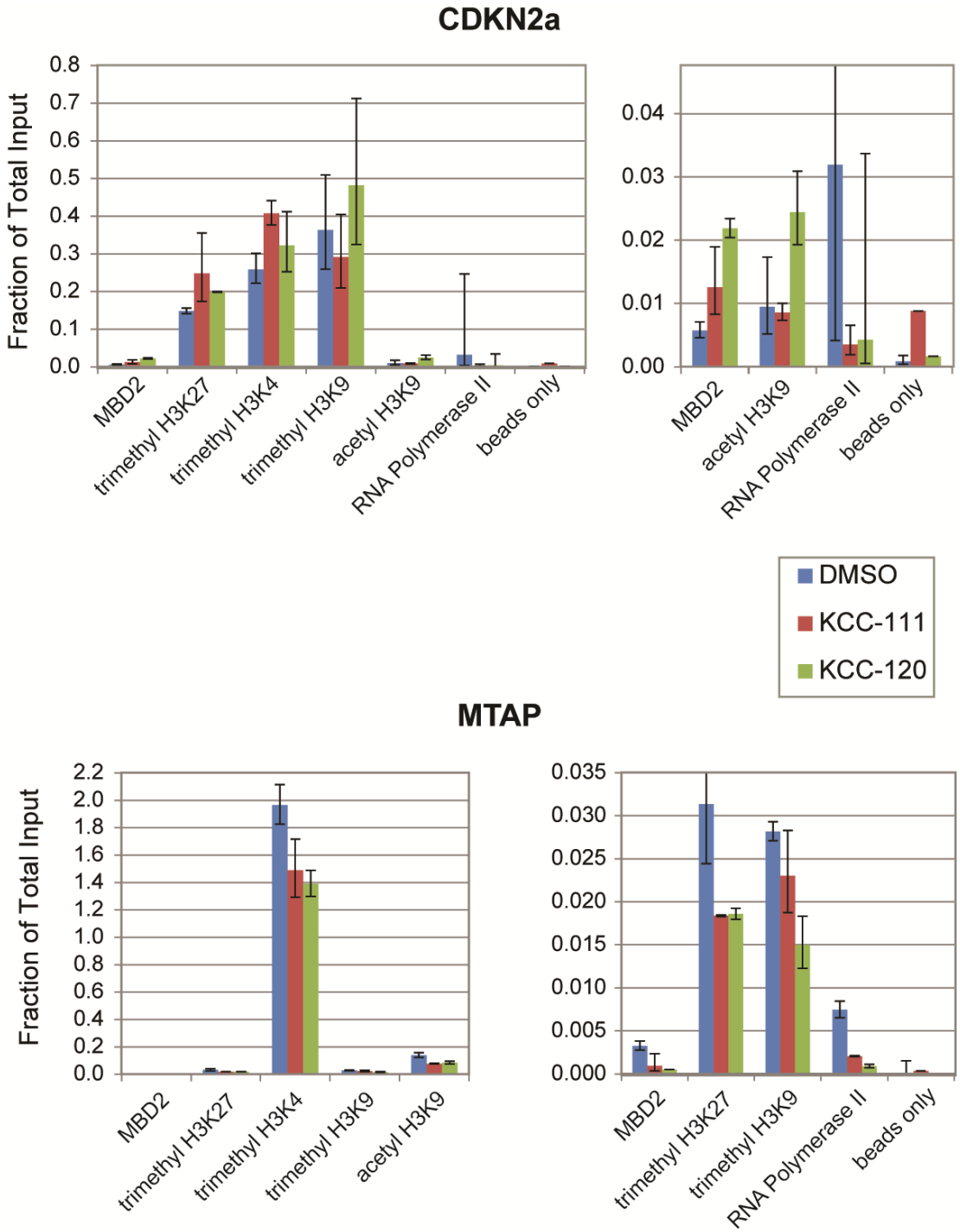


Figure 4.23: ChIP results of non-epigenetically controlled genes treated with KCC-120 and 111 cont...

Fraction of input for genes with unmethylated promoters and low RNA expression. RNA polymerase and MBD2 precipitations had universally low results. Bar plots on left and right for each gene are the same data plotted with different y axis scales. MTAP results are consistent with a gene that is highly expressed, contradictory to what is published in the literature.

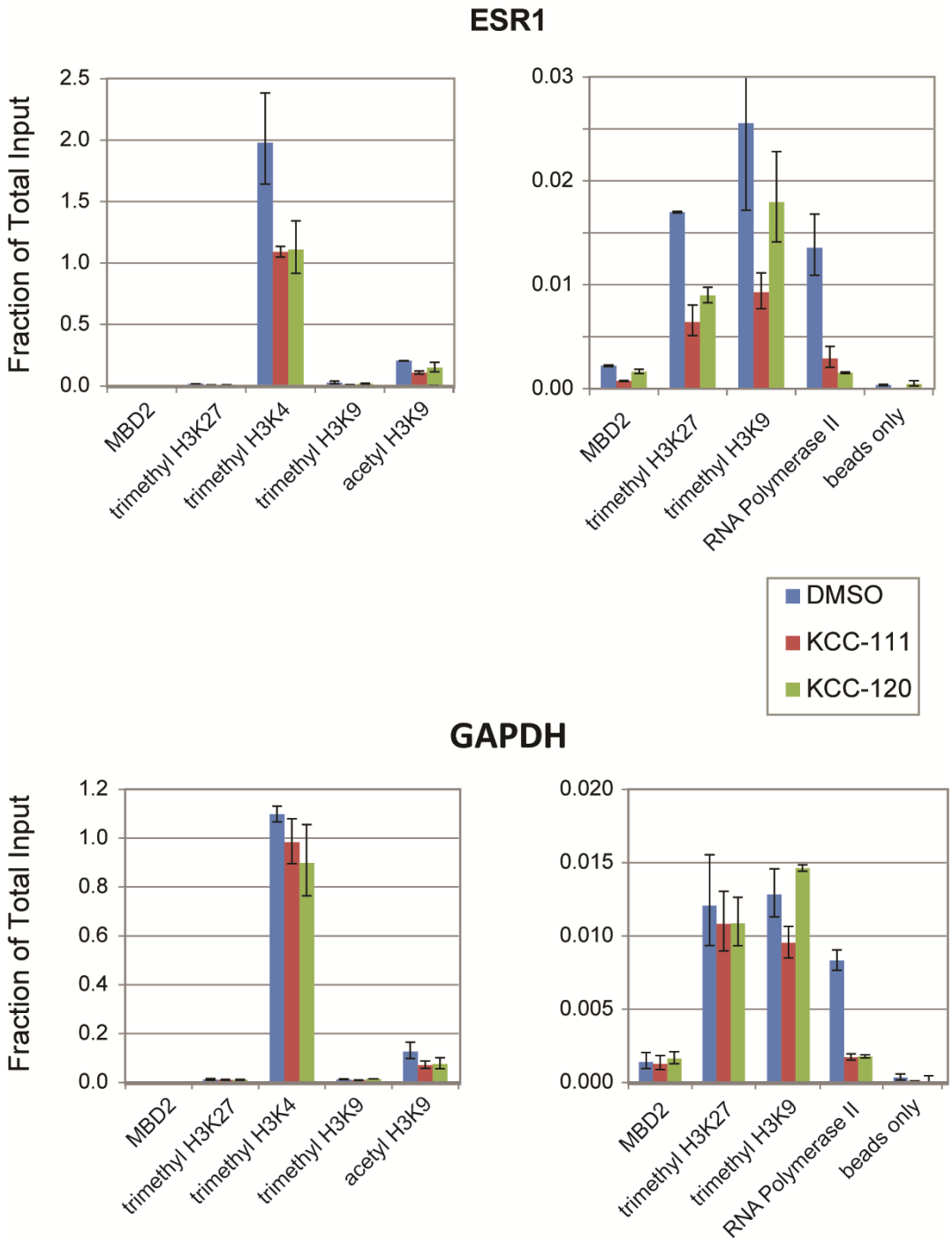


Figure 4.24: ChIP results of highly expressed genes with KCC-120 and 111

Fraction of input for genes with unmethylated promoters and high RNA expression. RNA polymerase and MBD2 precipitations had universally low results. Bar plots on left and right for each gene are the same data plotted with different y axis scales. Both genes have chromatin significantly enriched for active marks.

KCC-111

	Methylated Promoter	RNA Expression	H3K27 trimeth	H3K4 trimeth	H3K9 trimeth	H3K9 acetyl	MBD2	Histone Change
GSTP1	Yes	Low	same	same	same	up	low	On
DAPK1	Yes	Low	down	up	same	up	down	On
KLK10	Yes	Low	down	up	down	up	low	On
TWIST_1	Yes	Low	down	same	same	down	down	Amiguous
RAR Alpha	Semi	Low	same	down	same	down	down	Off
TWIST_3	No	Low	down	same	up	down	down	Off
CDKN2a	No	Low	up	up	same	same	up	Ambigous
MTAP	No	Low	down	down	same	down	low	Off
GAPDH	No	High	same	same	down	down	same	Ambiguous
ESR1	No	High	down	down	down	down	low	Ambiguous

KCC-120

	Methylated Promoter	RNA Expression	H3K27 trimeth	H3K4 trimeth	H3K9 trimeth	H3K9 acetyl	MBD2	
GSTP1	Yes	Low	same	up	up	up	low	On
DAPK1	Yes	Low	same	up	same	up	same	On
KLK10	Yes	Low	down	up	same	up	low	On
TWIST_1	Yes	Low	down	up	same	down	down	On
RAR Alpha	Semi	Low	same	same	same	same	same	Ambigous
TWIST_3	No	Low	same	up	up	same	down	Ambigous
CDKN2a	No	Low	up	same	same	up	up	Ambigous
MTAP	No	Low	down	down	down	same	low	On
GAPDH	No	High	same	same	same	same	same	Ambiguous
ESR1	No	High	down	down	same	same	down	Ambiguous

Promotes Open Chromatin
No significant change
Promotes Closed Chromatin

Table 4.4: Summary of chromatin mark changes based on ChIP analysis of MCF7 cells treated with biologically active TR FRET assay inhibitors

4.4 Discussion

A key focus of our lab for a number of years now has been identification of novel small molecules to target the epigenetic reader protein MBD2. Previous students and fellows focused on cell based re-expression approaches, similar to the luciferase assay utilized in the work described here with the GSTP1 promoter. These methods are very effective at finding biologically active compounds, as hits from cell based screens must permeate the cell and cause a gene re-expression to be found. Unfortunately, finding the target of such small molecules is not always trivial, as they could be disrupting any protein in the silencing pathway. If one considers that the Mi-2 NuRD complex alone contains at least 18 identified components, and these proteins can interact with others directly or indirectly, the network of possible targets becomes quite large.

In this work we take an orthogonal approach and choose the target, in this case MBD2, and then determine which small molecules which inhibit MBD2 *in vitro* have biological consequences. After a rigorous effort we uncovered 2 compounds capable of relieving epigenetic repression. To our surprise they targeted DNA and not the protein, and were not specific for DNA methylation status. While toxic to all cells above 25μM, below this the compounds seem to induce histone changes which confer a more open chromatin state at only methylated gene promoters. The fact that the activity of these compounds is focused

towards methylated promoters, but not unmethylated or highly expressed promoters, suggests that their mechanism is likely due to disruption of epigenetic repression machinery. This observation also rules out nonspecific steric hindrance as a mechanism, as the increased RNA expression levels at these locations is consistent with the notion that transcription factors and RNA polymerases are still accessing the gene.

These results, coupled with the evidence they relieve epigenetic repression and induce gene expression, suggest that KCC-120 and 111 could hold value as epigenetic targeting agents. A cursory review of the literature shows small molecule interaction with DNA is not a novel epigenetic target mechanism. All current DNMT1 inhibitors are nucleoside mimics, and other groups continue to develop DNA targeting agents to disrupt other important cancer related proteins (*141, 162*). Whether the compounds identified in this screen are viable as potential drug candidates will require a significant effort to establish safety and tolerable toxicity profiles in both cell based and animal models.

Three other compounds we investigated, KCC-131, KCC-107 and KCC-127, remain somewhat difficult to judge. Both clearly inhibit our TR FRET based screen, disrupting meDNA-MBD2 interaction. None of the follow-up experiments presented here, however, explain the mechanism behind the disruption. Since none showed biological activity it would be easy to waive them off as false positives. But if one considers the

numerous potential reasons why a small molecule may not access its site of action in a cell based screen they are harder to dismiss. The evidence here suggests that they are likely weakly bound to either DNA or protein. Parsing out which will require more sensitive techniques, such as surface plasmon resonance or some other novel technology. While this may seem like a fruitless venture, even weakly bound small molecules can be curated into useful inhibitors in the age of structure aided design. With the crystal structure of MBD2-MBD likely to be solved by our collaborators in the near future, further investigation into these compounds is warranted.

One of the most surprising outcomes of this effort were the results of NF449. This small molecule, identified in our LOPAC$^{1280™}$ screen, is not really that small, weighing in at 1505 Daltons. A brief search on pubchem shows it as a frequent hit on many small molecule screens, 32 hits on 174 total assays or 18.4%, particularly those utilizing DNA binding proteins or cell surface receptors (*163, 164*). The eight negative charges it contains likely contribute to this finding, allowing NF449 to make numerous favorable interactions with positively charged side chains on a protein's surface. Despite literature suggesting its activity is relatively non-specific, we observed reproducible, dose dependent binding by ITC and TR FRET assays to MBD2-MBD. We can infer, then, that the small molecule binds in a specific manner at a targeted site(s).

Very surprisingly, a look at the structure of NF449 shows it contains numerous carboxyl like structures. When this is considered with the increased melting stability of MBD2 in the presence of formate, and that the first condition which yielded diffraction worthy crystals was citrate, both carboxyl group containing salts, it should not be surprising that a carboxyl like moiety is present in an inhibitor. Given its size and nanomolar IC_{50} for MBD2, it is doubtful the entire NF449 molecule is interacting with the protein. It would be interesting, however, to determine which part of it is contacting MBD2. This information could allow for design of a more specific small molecule. Condition screening for co-crystallization of MBD2 and NF449 should also be considered, since carboxyl groups have made favorable conditions for previous crystallography attempts. This is an especially attractive opportunity as co-crystallization of MBD2 and NF449 would allow for structure aided design, and supply a DNA free structure of the protein: two birds with one stone made of x-rays.

Despite not discovering an immediately attractive nanomolar small molecule inhibitor for MBD2 our HTS effort yielded numerous interesting compounds to follow up. With both DNA and protein targeting compounds the future is bright for finding a novel class of epigenetic reader protein disruptors.

4.5 Methods

Primary Screening for MBD2(MBD)-meDNA Interaction Inhibitors

The TR FRET assay was performed as described in chapter 3, under the Time Resolved Fluorescence Resonance Energy Transfer (TR-FRET) heading. The primary screen was run using 4.4μM inhibitor in 10% DMSO in 1536 well plates with a working volume of 5μL. All plates were run by The Scripps Research Institute® in Florida.

UHRF1_SRA protein Purification

A codon-optimized sequence of the SRA domain of UHRF1 was synthesized in a pUC19 plasmid (GenScript USA Inc), and cloned into the HIS-tag containing vector pEXP-CT (life Technologies). This vector was then transformed into BL-21 Gold cells (Agilent) and cultured at 37C and 220rpm in an orbital shaker to log phase (OD_{600} of 1.0) in Terrific Broth media. Cultures were then induced for protein expression using 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG, Corning Cellgro) at 24C overnight. Bacterial cultures were harvested by centrifugation at 3,000 x *g* for 15 minutes at 4C. Isolated bacterial cells were resuspended in an equilibrium buffer consisting of 300 mM NaCl, 50 mM sodium phosphate pH 8.0, 5 mM imidazole (Sigma Aldrich), and a protease inhibitor cocktail (Roche USA). Lysis was performed by adding 0.75mg/mL lysozyme followed by incubation for 30 minutes at room temperature. Additionally,

the bacteria mixture was sonicated three times (Sonic Dismembrator Model 100, Thermo Fisher Scientific, Waltham, MA) for 10 seconds followed by 30 seconds on ice. The mixture was centrifuged at 10,000 x *g* for 30 minutes to remove non-soluble components. IMAC Nickel NTA beads (Bio-Rad) beads were prepared by washing with deionized water followed by equilibrium buffer. The bacterial supernatant was mixed with washed beads and incubated at 4C for 1 hour. The beads were collected on a filter column and washed with 5 solutions of increasing imidazole concentration (15, 20, 25, 30, 45 mM). The beads were then eluted in equilibrium buffer containing 250 mM imidazole and incubated at 4C for 20 minutes. The eluent was collected by gravity flow and dialyzed into a protein storage buffer containing 125mM NaCl, 20 mM Tris pH 7.4, and 10 mM 2-mercapto ethanol using the Slide-A-Lyzer system (Pierce).

TR-FRET assay for UHRF1_SRA binding to methylated DNA

TR-FRET assays were performed in an optimized buffer (4% glycerol, 1 mM MgCl2, 0.5 mM EDTA, 0.5 mM DTT, 125 mM NaCl, 10 mM Tris HCl, and 0.2% Tween-20) in non-binding, flat-bottom 384-well plates (Corning) at a total volume of 20µL. To make a master mix, UHRF1_SRA protein was added at 125nM and terbium-labeled antibody was added at 5nM (life Technologies). For binding verification experiments, FAM-labeled oligonucleotides were added in a dilution series, and the reaction mixture was incubated at 4C with shaking for 1

hour. FRET measurements were made on a SpetraMax M5 (Molecular Devices), using a excitation of 332 nm and emission read at 515 nm with a 50μsec delay and 400μsec integration.

Dose Response Series on Potential MBD2-MBD Inhibitors

Dose response curves on the 46 available compounds were performed using the TR FRET assay described above in the primary screen. Experiments were performed in 1536 well plates and a working volume of 5μL. Compounds were run in triplicate from ~10nM to 100μM. The assay, data collection and fitting was performed in collaboration with The Scripps Research Institute® in Florida.

Recombinant protein for MBD family members as made in BL21-DE3 cells (Agilent). Bacterial codon optimized DNA sequence was obtained in a pUC19 vector (GenScript) and cloned into a pEXP-5CT C-terminal his tag labeled bacterial expression vector per manufacturer's instructions (Life Technologies). All proteins were expressed in 1L of terrific broth (Corning-Cellgro) with 1mM IPTG overnight at 20°C as previously described(*110*). Protein was purified using affinity purification as previously described(*110*) and concentrated to ~100μM (~1mg/mL). SP-1 protein (ab82236) was purchased from Abcam and resuspended to 5μM in 150mM NaCl, 10mM Tris, 1mM TCEP and stored at -80°C.

Dose response curves for the MBD family members MBD1-MBD, MeCP2-MBD, SP1, MBD2-MBD and UHRF1-SRA were performed in 384

well plates as previously described(*110*) with the following changes: Dose response concentrations began at 2000nM and proceeded in a 4 fold dilution to 0.12nM in quadruplicates. SP-1, MBD1-MBD and MeCP2-MBD protein were used at 10nM, MBD2-MBD protein at 25nM, and UHRF1-SRA protein at 125nM final concentration in a working volume of 20µL.

Epigenetically Silenced Gene Re-expression Potential of MBD2 Active Compounds

Endogenous Gene Expression Analysis

MCF-7 cells (ATCC) were grown in MEM media (Life Technologies) with 10% fetal bovine serum (Gemini Biosciences). LNCaP cells (ATCC) were grown in RPMI media (Life Technologies) with 10% fetal bovine serum (Gemini Biosciences). Cells were dosed with either 50µM (MCF7) or 20µM (LNCaP) of each compound in 0.5% DMSO (Sigma Aldrich) for 48 hours. Cells were harvested and RNA was extracted via an RNeasy kit (Qiagen). RNA was converted to cDNA with the High Capacity cDNA Kit (Applied Biosystems) and real time PCR was performed on the epigenetically silenced gene Glutathione S Transferase Pi 1 (GSTP1) using the primers 5′-GGGACCCTCCAGAAGAGC and 5′-ACTCACTGGTGGCGAAGACT, and Cox 2 (PTGS2) as previously described(*63*) on a CFX 96 machine (Bio-Rad). Graphs were constructed in excel (Microsoft).

Dose response expression was obtained on MCF7 cells cultured as described above. Cells were treated for 48 hours with 25μM, 12.5μM, 6.3μM, or 3.1μM compound. RNA was extracted and real time PCR performed as described above. Primers used for this assay included:
DAPK1: Forward 5'-CTCACGGCATTTCTTCACAA-3', reverse 5'-TGGTGAGGCGTGACAGTTTA-3'.
KLK10: Forward 5'- CCGTTGAAGAGCGAGACCT-3', reverse 5'-CTGCTGATGGCGCAACT-3'.

Luciferase Expression Analysis

A ~400bp GSTP1 promoter driving luciferase in plasmid pGL3 (Promega) was acquired from previous lab members. A CMV promoter driving renilla was obtained from Promega. 200ng of the pGL3 plasmid and 50ng of the CMV-renilla plasmid were transfected into HEK293 (ATCC) cells grown in DMEM (Life Technologies) with 10% FBS (Gemini Biosciences) using MegaTran 1.0 (OriGene) per manufacturer's instructions in 24-well culture plates (Corning) with biological duplicates. 8 hours after transfection cells were treated with compound at 50μM, 20μM, or 5μM. The next morning the media was replaced and compounds reapplied. After 48hrs total exposure to small molecule cells were lysed and read for luciferase using a Dual-Luciferase® Reporter Assay (Promega) following manufacturer's instructions. P-values were

determined by student t-test using the open source software R version 3.0.2 (*165*)

.

Biophysical Interactions of Active Compounds with meDNA and MBD2 Protein

DNA Intercalator Competition Assay

The GSTP1 promoter was PCR amplified from the pGL3 (Promega) plasmid used for the luciferase expression experiment described above. PCR was performed with Platinum™ Taq (Life Technologies) per the manufacturer's instructions with forward primer GGGACCCTCCAGAAGAGC and reverse primer AGTGCCGTTAGCGGCTTTC, at 58^{o}C annealing temperature and with 40 cycles. PCR product was run on a 1% agarose gel and purified with a Gel Purification Kit (Qiagen). 10ng of eluted product was cloned into a topo 2.1 vector (Life Technologies) per manufacturer's instructions. This new plasmid was used untreated (unmethylated) or treated with M.Sssi enzyme (New England Biolabs) per manufacturer's instructions (methylated). Methylation was confirmed by digesting the M.Sssi plasmid with HhaI (New England Biolabs) and HpaII (New England Biolabs) restriction enzymes per manufacturer's instructions.

For the intercalator experiment, 100ng of plasmid (either methylated or unmethylated) was mixed with 10x Sybr® Reagent (Bio-

Rad). Compound was added at 500μM, 50μM, 5μM or 0.5μM in a final working volume of 20μL in 384 black well low volume plates (Corning). Triplicates were performed for each dose. Plates were incubated at room temperature with gentle shaking and read on a Spectramax™ M5 plate reader with excitation at 488nm and emission read at 520nm. Data was analyzed using the open source software R version 3.0.2 (*165*).

Isothermal Calorimetry Analysis of MBD2-meDNA Active Compounds

MBD2-MBD protein, purified as described earlier, was dialyzed overnight in 3mL 3000MWCO dialysis cassettes (Thermo Scientific-Pierce Net) after they were pre-soaked in buffer for 5 minutes. Protein was dialyzed overnight at 4^{o}C with gentle mixing by stir bar in 500mL of ITC buffer (125mM NaCl, 1mM MgCl2, 0.5mM EDTA, 0.5mM TCEP, 10mM Tris pH 7.4, all from Sigma Aldrich). The following day protein was concentration was verified by NanoDrop(Thermo Scientific) to be 0.5-1mM and flash frozen in liquid nitrogen for storage. For ITC experiments the calorimeter was loaded with 2mL of either 50μM MBD2-MBD in 5% DMSO or 30μM singly symmetrically methylated hairpin oligo (ATGCT^{me}CGTAGCACTTTTGTGCTA^{me}CGAGCAT) (IDT, HPLC purified) in 5% DMSO, depending on whether protein or DNA was being tested. The syringe was loaded with compound at 5% DMSO as indicated in table 4.5, varying between 750μM and 1.5mM. Data was analyzed and fitted

using the Origin® software provided by MicroCal, utilizing a chi squared statistic to determine optimal fitting curves.

Chromatin Immunoprecipitation (ChIP) of Histone Modifications at Epigenetically Silenced Regions

ChIP experiments were performed on MCF7 cells (ATCC) at passage number ~10 grown in MEM media (Life Technologies) supplemented with

Cell (2000µL)	Syringe (500µL)	Compound Tested	Temp (C)	Notes
20µM MBD2-MBD	100µM meDNA	meDNA	25	
30µM meDNA	750µM	KCC-101	25	
30µM meDNA	750µM	KCC-102	25	
30µM MeDNA	750µM	KCC-107	25	
75µM MBD2-MBD	1.5mM	KCC-107	25	
50µM KCC-111	225µM meDNA	KCC-111	25	
30µM meDNA	750µM	KCC-116	25	
50µM MBD2-MBD	1mM	KCC-116	25	
30µM meDNA	750µM	KCC-120	25	
30µM meDNA	750µM	KCC-127	25	Insoluble at 50µM, ran with some precipitant present
50µM KCC-131	225µM meDNA	KCC-131	25	Insoluble above 100µM
50µM MBD2-MBD	750µM	KCC-131	25	Insoluble above 100µM
50µM MBD2-MBD	750µM	NF449	25	
30µM meDNA	750µM	NF449	25	
Buffer only	750µM	NF449	25	

Table 4.5: Setup for isothermal calorimetry experiments with HTS active compounds

Unless compound was insoluble at high concentration, small molecule was injected into meDNA or protein.

10% fetal bovine serum (Gemini Biosciences). Compounds were diluted in DMSO to 10mM and eight T-175 flasks (Sarstedt) of cells were treated with 15μM KCC-111 (in 0.5% DMSO), 20μM KCC-120 (in 0.5% DMSO) or 0.5% DMSO (control) for 24 hours in a 37°C incubator and 5% CO_2. After 24 hours cells were washed with PBS (CellGro) and fixed with 20mL of freshly mixed 1% formaldehyde (Sigma Aldrich) in PBS for 10 minutes at room temperature on a shaker. Fixation was stopped by adding 1mL of 2.5M glycine and incubating for 5 minutes on a shaker at room temperature. The formaldehyde/glycine mixture was aspirated and cells were washed with 10mL cold PBS, scraped into 5mL of cold PBS, and centrifuged at 1000xg to pellet. Supernatant was removed and pellets were resuspended in 8mL of cold PBS, aliquoted equally into 8 Eppendorf tubes (USA Scientific), re-pelleted as described above and flash frozen in liquid nitrogen and stored at -80°C. ChIP was performed by the wonderful Lord Esopi following the protocol listed below, kindly provided by Dr. Michael "the Haff" Haffner. Buffers for the protocol below can be found in table 4.6.

ChIP PROTOCOL(*166*):

1. Defrost cell pellets on ice.
2. Lyse cell pellet in Lysis Buffer supplemented with protease inhibitor. Use 350μl for per T175 flask of 80% confluent LAPC4 or LNCaP.
3. Sonicate to a fragment size of 300 – 800bp using the Covaris sonicator in NGSC.

4. Set aside 10µl of the sonicated sample material, label as “input” and freeze at -20°C.
5. Dilute the remaining sample with dilution supplemented with 1 X Roche protease inhibitor cocktail EDTA free buffer 1:4.
6. Prepare Dynabeads protein G (Life technologies) by washing 60 µl/sample of resuspended bead solution twice with PBS supplemented with BSA (5mg/ml). Use a magnetic rack for all wash steps.
7. Pre-clear samples by incubation the diluted sheared material with 30 µl bead for 1 h on a spinning wheel in the cold room.
8. Apply tubes to the magnetic rack, let beads precipitate and transfer samples in new pre-chilled tubes. Discard beads.
9. Add 5µg of appropriate antibody listed below to the sample and incubate over-night in the cold room.
   - MDB2: Bethyl A301-633A
   - H3K27 trimethyl: Millipore 07-449
   - H3K4 trimethyl: Abcam ab-8580
   - H3K9 trimethyl: ab-8898
   - H3K9 acetyl: Millipore 07-352
   - RNA Pol II: Covance MMS-126R-500
10. Block 30 µl of beads for each sample by incubating in PBS BSA (5mg/ml) supplemented with 100 µg/sample of yeast tRNA over-night in the cold room.
11. Next day, add 30 µl of blocked bead solution to the samples and

incubate for 4 h.

12. Apply tubes for 30 sec to magnet to precipitate bead bound immunocomplexes.
13. Sequentially wash beads for 10 min on ice each with TSE1, TSE2, TSE3 and TE (pH. 8.0).
14. After the final wash step, re-suspend beads in 100 µl freshly prepared IP Elution Buffer. Vortex to mix and transfer bead mixture into strip tubes.
15. Incubate at 56°C for 15 min, vortex every 5 min, to elute the chromatin from the beads.
16. Collect the supernatant, which contains the DNA. Repeat this step again by eluting with another vol. of 100 µl IP Elution Buffer for 15 min.
17. Incubate the supernatant containing the DNA at 65°C for 8 h to complete the reversal of the formaldehyde cross-links.
18. *Do not forget to de-crosslink the input faction*
19. After de-crosslinking, add 8 µl of RNaseA (10 mg/ml) to each sample and mix by inverting tube several times and incubate at 37°C for 2 hours.
20. Add 4 µl of Proteinase K (20 mg/ml) to each sample and mix by inverting tube several times and incubate at 56°C for 2 hours.
21. Purify DNA using a Qiagen PCR purification kit, elute sample with 100 µl water.

After extraction real time PCR was performed in duplicates to determine enrichment. Briefly, wells were mixed with 10µL of IQ™ Sybr® Green Master Mix (Bio Rad), 2µL eluted DNA, forward and reverse primers to a final concentration of 400nM (IDT), and water up to a final volume of 20µL. Primers used for this analysis are listed in table 4.7. Analysis was performed on results in excel (Microsoft) using a pseudo $\Delta\Delta C_q$ analysis: C_q values for each treatment were compared to both their total input fraction and DMSO to determine a fraction of input. Results were plotted against each other with standard deviation show by error bars.

Lysis Buffer	*1% SDS, 5 mM EDTA, 50 mM Tris.HCl [pH 8.1]*		
			for 100ml
	1% SDS		1 g
	5mM EDTA	FW = 372.2	186mg
	50mM TRIS.HCl pH 8	Stock 1M	5ml
	protease inhibitors	50x	

Dilution Buffer	*1% Triton-X100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris.HCl [pH 8.1]*		
			for 100ml
	1% Triton-X100		1 ml
	2 mM EDTA	FW = 372.2	74.4 mg
	150 mM NaCl	FW = 58.44	876.6 mg
	20 mM TRIS.HCl		2 ml
	protease inhibitors	50x	

TSE I	*0.1% SDS, 1% Triton X*		
			for 100ml
	0.1% SDS		100mg
	1% Triton		1 ml
	2 mM EDTA	FW = 372.2	74.4 mg
	20 mM Tris.HCl		2 ml
	150 mM NaCl	FW = 58.44	876.6 mg

TSE II	*0.1% SDS, 1% Triton X 100, 2 mM EDTA, 20 mM Tris.HCl, pH 8.1, 500 mM NaCl*		
			for 100ml
	0.1% SDS		100 mg
	1% Triton		1 ml
	2 mM EDTA	FW = 372.2	74.4 mg
	20 mM Tris.HCl		2 ml
	500 mM NaCl	FW = 58.44	2.9 g

TSE III	*0.25 M LiCl, 1% NP 40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris.HCl, pH 8.1*		
			for 100ml
	0.25 M LiCl	FW = 42.39	1.06 g
	1% Igepal-CA630		1 g
	1% deoxycholate		1 g
	1 mM EDTA	FW = 372.2	37.2 mg
	10 mM Tris.HCl		1 ml

IP Elution Buffer	*1% SDS, 0.1M NaHCO3, proteinase K*
	for 10 ml
10% SDS	1 ml
1M sodium bicarbonate	1 ml
Proteinase K (20mg/ml)	*add 2 ul directly to sample*

Table 4.6: Buffers for ChIP protocol

Target	Forward Sequence	Reverse Sequence	Efficiency %	Melt Temp (C)	Product (bp)
GSTP1	GGGACCCTCCAGAAGAGC	ACTCACTGGTGGCGAAGACT			134
DAPK1_1	CTTGCAGGGTCCCCATTG	GTCCGGCTGTCCTCCTCA	87.4	88.5	98
KLK10_1	GAGCTCTCATGGCCAGGATC	CAGGTCAAAAGGGTGCTTGC	93.11	88.0	135
TWIST1_1	TCCTGGGCGTTTCTGAAGAC	TGGGTCGTTGTAGAGGGGAA	86.79	88.0	106
TWIST1_3	GGAAATCGAGGTGGACTGGG	GGGCCGGAGACCTAGGTAA	77.37	89.0	103
RARA_1	CCCTCTGCCTGTGTTTGTG	CGGCTTGTGTCATTGTCTGT			96
GAPDH_2	TCCAATTCCCCATCTCAGTC	GCAGCAGGACACTAGGGAGT			91
ESR1_1	CTCGGGCTGTGCTCTTTTTC	CCAGATGCTTTGGTGTGGAG	97.82	88.5	125
CDKN2a_2	GAATGTGGCACCCCTGAAGT	CAGTCCTCCTTCCTTGCCAA	88.37	88.0	101
MTAP_2	GGTCTCAACCCCCTAAAAGC	AGGGATGACGAAAGGGAAAT			115

Table 4.7: Table of primers for ChIP real time PCR analysis

4.6 Acknowledgements

I would like to acknowledge the Scripps Institute® of Florida, in particular Virneliz Fernandez-Vega, Tom Bannister and Peter Hodder for their help planning and executing the high throughput screen. Without their advice and facilities this project would never have been possible. I would also like to acknowledge Hugh Giovinazzo and Dave Walker for their help with all of the NF449 ITC data and assay development respectively. David Esopi's tireless efforts performing all of the ChIP experiments described here were invaluable, and it was a true pleasure collaborating with such a hardworking and smart individual. And to Micheal Haffner, for his invaluable advice and discussion regarding the project, in particular the ChIP experiments, thank you. He also provided the ChIP protocol, which was awesome.

This work was supported by NCI grant R03MH098712, the Prostate Cancer Foundation grant, the Commonwealth Foundation, and Department of Defense Prostate Cancer Research Program Pre-doctoral training grant (W81XWH-11-1-0618; to NW).

Chapter 5

Investigation of Whole Genome Methylation Patterns as a Risk Stratification Tool in Prostate Cancer

5.1 Abstract

Recent evidence suggests that a significant fraction of men with prostate cancer undergo unnecessary and life altering therapy. Accurate prognosis with pathology is possible if good sampling of the tumor is accomplished. This is limited, however, by lack of suitable radiographic methods for targeting biopsies. Here we describe an investigation to determine whether DNA methylation could assist with Gleason scoring or act as an independent predictor of prognosis.

We determined whole genome methylation levels using Illumina 450K Methylation Microarrays and a novel MBD2-MBD protein based methylated DNA enrichment technique coupled with next generation sequencing. Following a thorough pathology re-review and quality control, we obtained 72 samples for this analysis and binned them into four groups: normal prostate, low grade (Gleason score ≤6), high grade (Gleason score ≥8) and metastasis. Comparing peaks of methylation determined by our novel MBD-Seq method and the microarray beta values, we found a strong correlation between the two platforms.

We next used the microarray beta values to determine the utility of DNA methylation marks for predicting Gleason score. After removing probes with CpG densities <0.025 and one sample which had its tumor and normal samples switched, we calculated the standard deviation across all samples for each probe. The top 300, 1000 and 20,000 (5%) of total probes based on standard deviation all showed nice separation

between low grade and normal prostate tissue versus high grade and metastasis by principle components analysis.

We queried the top 5% most variable probes for hypermethylated regions in only high grade and metastatic samples with a sensitivity of ≥0.25, and specificity ≥0.9. This yielded 72 probes, 53 genes, which comprised a possible biomarker panel predicting Gleason score. Visualization of the probes and samples by heat map and application of a Tukey Honest Significant Difference and ANOVA statistic showed that the panel could significantly predict Gleason score/tumor stage based on methylation values with very significant p-values <2.2e-16. Gene ontology analysis did not yield any enriched gene signatures. The panel performed best when differentiating between metastatic and primary cancer or normal prostate tissue, with a difference in methylation levels >31%.

Here we describe application of a methylated DNA enrichment technique and confirm that it can accurately determine methylation levels when compared to established methylation microarray technology. We also show that DNA methylation could serve as a useful biomarker to assist with Gleason scoring and may even act as an independent, or improved, method for risk stratification in the biopsy setting.

5.2 Introduction

Despite many advances prostate cancer remains one the chief causes of cancer related morbidity and mortality for men today. Yet, autopsy studies have shown that a significant fraction of men will ultimately die with but not from the disease. Therefore, a major controversy has arisen regarding the potential overtreatment of prostate cancer with some studies suggesting that under current clinical practices, it is necessary to treat between 5 to 45 men with potentially harmful medical procedures in order to prevent a single death from prostate cancer(*167-170*). The heterogeneity of the disease is quite remarkable considering that some men will die of recurrent disease despite aggressive treatment including surgery or radiation while others live with the disease for decades without major morbidity. This complexity highlights the need to develop biomarkers that can distinguish the men with aggressive disease in order to prevent subjecting men with indolent forms of cancer to unnecessary treatments that carry potentially life-altering side-effects, such as incontinence and impotence(*171*).

The greatest utility for prognostic biomarkers occurs in the period of initial screening and diagnosis. In this setting molecular biomarkers could assist with both the diagnosis of cancer as well as the identification of those most likely to behave aggressively. This dual approach would help efficiently and correctly allocate aggressive

treatment options, such as prostatectomy or radiation therapy, to patients with aggressive forms of the disease while saving conservative approaches, such as watchful waiting, for those with more indolent disease.

Gleason score is an extremely effective histopathological biomarker for assessing the aggressiveness of prostate cancer in the biopsy setting. However, it is limited by several factors. First, Gleason score must be assessed in tissue specimens and thus cannot be used in the initial screening or diagnostic setting until after a biopsy has been completed. Since prostate biopsies are done without radiographic guidance with respect to tumor location, sampling issues can prevent proper assessment of Gleason score. This has been demonstrated by the fact that a significant portion of men are "up-graded" with higher Gleason scores when they have their radical prostatectomies(*172*). Second, Gleason scores are a somewhat subjective measure, and much literature has shown that there is significant inter-observer variability in the assessment of Gleason score even by expert pathologists(*172*). We hypothesize that DNA methylation changes associated with Gleason grade can be exploited to help reduce the variability associated with Gleason scoring for more robust prostate cancer risk stratification.

The progressive acquisition of somatic genome alterations is a defining feature of all human cancers, including prostate cancer. It is well established that genetic alterations such as somatic mutations,

amplifications, translocations, and deletions are central to the development and progression of the disease (*173*). More recently, research has implicated epigenetic processes, such as alterations in DNA methylation patterns, as major driving forces of prostate cancer initiation and progression (*134, 135*). Epigenetic changes such as cytosine methylation or histone tail modification can activate proto-oncogenes or turn off tumor suppressors (*134, 174*). These changes ultimately yield phenotypes resembling those seen with genomic landscape changes: either up-regulation of an oncogene or repression of a tumor suppressor (*175, 176*). For prostate cancer, DNA cytosine hypermethylation, capable of stably silencing gene expression, appears to occur in multiple waves (*134*). A large initial wave of CpG island hypermethylation takes place very early during prostate carcinogenesis; even arising at the stage of prostate precursor lesions and is maintained throughout disease progression (*134*). These early CpG island hypermethylation changes are already under large-scale clinical and translational development as biomarkers for prostate cancer screening and diagnosis.

Our laboratory and others have also collected preliminary evidence suggesting that there are subsequent waves of CpG island hypermethylation in prostate cancer (*134*). These changes may make excellent biomarkers for risk stratification as they should correlate with disease progression. Presumably these epigenetic alterations also play a role in driving disease evolution and should therefore be directly

associated with disease severity or aggressiveness (e.g., cancer grade and stage) and/or chance of recurrence after treatment depending on which changes are present in a particular cancer. Another potentially valuable outcome to understanding which genes are driving progression is identification of novel genes for targeted therapy. While this is not the primary focus of the project it would be an important discovery and could lead to new treatments.

Our lab has developed a novel technology called qMBD-Seq for observing genome-wide DNA methylation patterns (the methylome) at extremely high resolution (approaching 100bp). In preliminary studies, our group and others, using parallel approaches, applied this technology to analyze genome-wide DNA methylation patterns in a set of prostate cancer cell lines (*177*),(*178*). Comparing LNCaP prostate cancer cells to normal prostate epithelial cells (PrEC) we found the genes GSTP1 and PTGS2 clearly show an increase in methylation in their respective promoter regions for LNCaP but not PrEC. This is an expected result as both have been widely shown to be hypermethylated in LNCaP cells and in prostate cancer cells in human tissues, but almost never methylated in PrEC's or normal prostate tissues (*179, 180*). This analysis also yielded >400 novel genes with hypermethylated promoter regions. Bisulfite sequencing performed on some of these regions confirmed the results observed in the qMBD-Seq analysis. This illustrates the high accuracy and resolution of the method.

In this work we compare the methylation patterns found with our novel MBD-Seq approach to the well-established Illumina 450K Methylation microarray to further confirm the accuracy of our novel sequencing method. We then establish that the patterns obtained with the 450K array can differentiate between highly aggressive and more indolent prostate cancers as a tool to assist with accurate Gleason scoring and possibly act as an independent predictor.

5.3 Results

Sample Acquisition and Quality Control

Samples for this project were acquired from three sources, all within the Johns Hopkins School of Medicine or Hospital, and consisted of organ donor normal, primary and metastasis samples. All material was re-reviewed by a single expert uropathologist to confirm both Gleason score and other pathology (see methods). Once confirmed by pathology, samples underwent a rigorous three step quality control metric before being added to the study. First they were analyzed by NanoDrop to establish their purity spectroscopically. One sample had an $OD_{260/280}$<1.75 and was rejected. The second step was accurate quantitation of double stranded DNA concentration by a fluorescence dye assay (Qubit). Numerous samples showed a discrepancy between their fluorescence dye and NanoDrop concentration due to impurities unaccounted for by the NanoDrop assay. Additional material was requisitioned for these samples to ensure enough DNA was available for both sequencing and microarray analysis. Unfortunately there was not enough metastatic material available to request more and those samples were utilized for MBD-Seq only. In the final QC check each sample was run on an agarose gel and imaged with ethidium bromide. This served as a second verification of DNA quality and also provided an estimate of the average DNA size for each sample. All samples, with the exception of

metastasis 19332, contained high quality DNA of average size > 10kb (Figure 5.1). Since we acquired only a limited number of metastatic samples, and more were difficult to obtain, we opted to include 19332 in our study.

In total 62 samples were collected for sequencing (Table 5.1). Of the 22 normal samples, 7 originated from organ donors and the other 15 from matched normals from primary tissue collected for the study. For 450K methylation array analysis all 33 primary and normal samples were used, but there was only enough material from two metastasis samples (19329, 19332). The remaining microarray samples, 9 metastasis and 5 normals, were previously run on this array by other members of the lab and the data was recycled for this project (*135, 181*).

Library Preparation and Sequencing

For whole methylome analysis samples were submitted for MBD-Seq next generation sequencing analysis and processed as described in the methods. Briefly, the MBD-Seq approach was developed in our lab and utilizes the methylated DNA specific binding properties of the MBD2-MBD protein fixed to a bead to precipitate methylated DNA from a genomic fraction. It is similar to the approach described by Lee et al in 2010(*178*). Once methylated DNA fractions are bound to the protein, washes are performed and the bound DNA is eluted from the

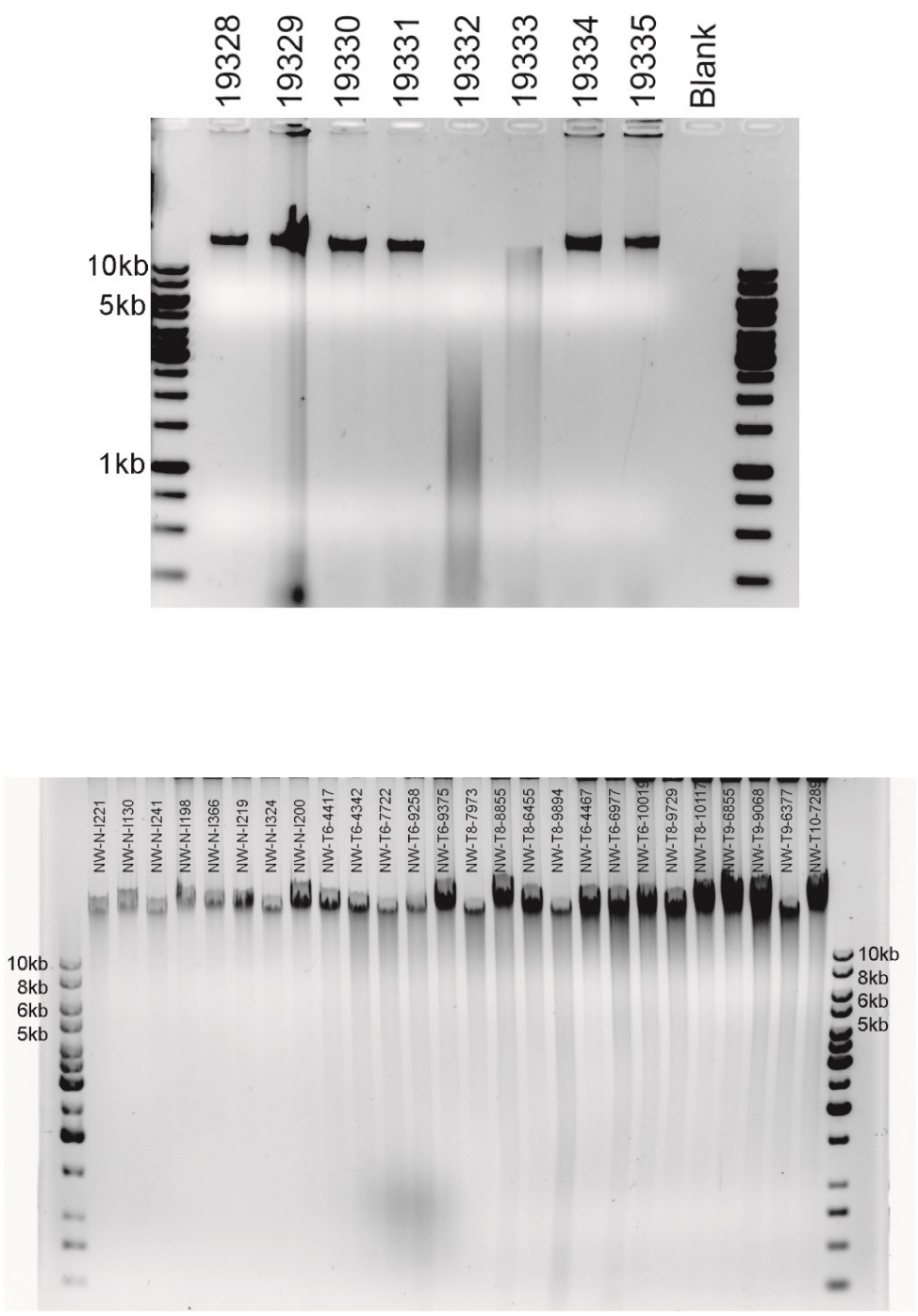


Figure 5.1: Representative agarose gel of samples acquired for MBD-Seq and microarray analysis.

	MBD-Seq	450K Methyl Array
Normal	22*	27*
Gleason 6	16	16
Gleason 8	7	7
Gleason 9	10	10
Metastasis	7	11**

*15 Matched Normals

**Only 2 samples overlapped with MBD-Seq

Table 5.1: Summary of samples acquired for MBD-Seq and microarray experiments.

protein by denaturation. The eluted product is then submitted for next generation sequencing library preparation (see methods). Un-enriched genomic DNA, taken prior to methylation specific precipitation, serves as a reference for each sample. Fully methylated lambda phage DNA was also spiked into each sample at a constant concentration. This was intended to serve as an internal standard to assess enrichment efficiency and to develop methods to measure the degree of methylation at each loci, simplifying inter sample comparison.

Despite our rigorous QC the initial two batches of samples, comprising 8 specimens, showed inconsistent DNA size distributions during library preparation when evaluated by Bioanalyzer. After an exhaustive investigation we suspect the samples may have contained some contaminant that was interfering with linker ligation or nick translation. All future samples were pre-treated with a repeat column purification to remove the contaminant post pathology review but pre QC. We encountered no further obstacles related to DNA quality.

450K Illumina Microarray analysis was performed on all of the normal and primary tissue as well as 2 of the metastatic samples. Raw data was obtained by submitting samples to the Sidney Kimmel Microarray Core Facility, but processed and analyzed by our group. As discussed above, previously acquired raw data was also included to increase the number of metastatic and normal samples.

Processing of MBD-Seq Raw Data

MBD-Seq data was processed and aligned to the hg18 genome as described in the methods. Post alignment analysis of the lambda phage sequences, which we intended to use as an internal standard for quantifying the level of methylation across the genome, showed fairly variable levels of spike-in. This was especially reflected in the enriched fractions (Figure 5.2). This unanticipated result complicated our initial approach for determining differentially methylated regions. A number of alternative strategies are being pursued to facilitate semi-quantitative differential methylation analysis.

Processing and QC of Illumina 450K Methylation Array Data

In parallel with the sequencing efforts we analyzed the methylation status of these same samples on the well-established Illumina 450K DNA Methylation Microarray. After obtaining raw fluorescence values, the data was normalized and processed as described in the methods using the Bioconductor–minfi Data package for R(*182*). Briefly, the probe type and color spaces were preprocessed and subset quantile normalization was performed to obtain beta values (fraction methylated) for each probe. The beta values attained with this package closely matched those

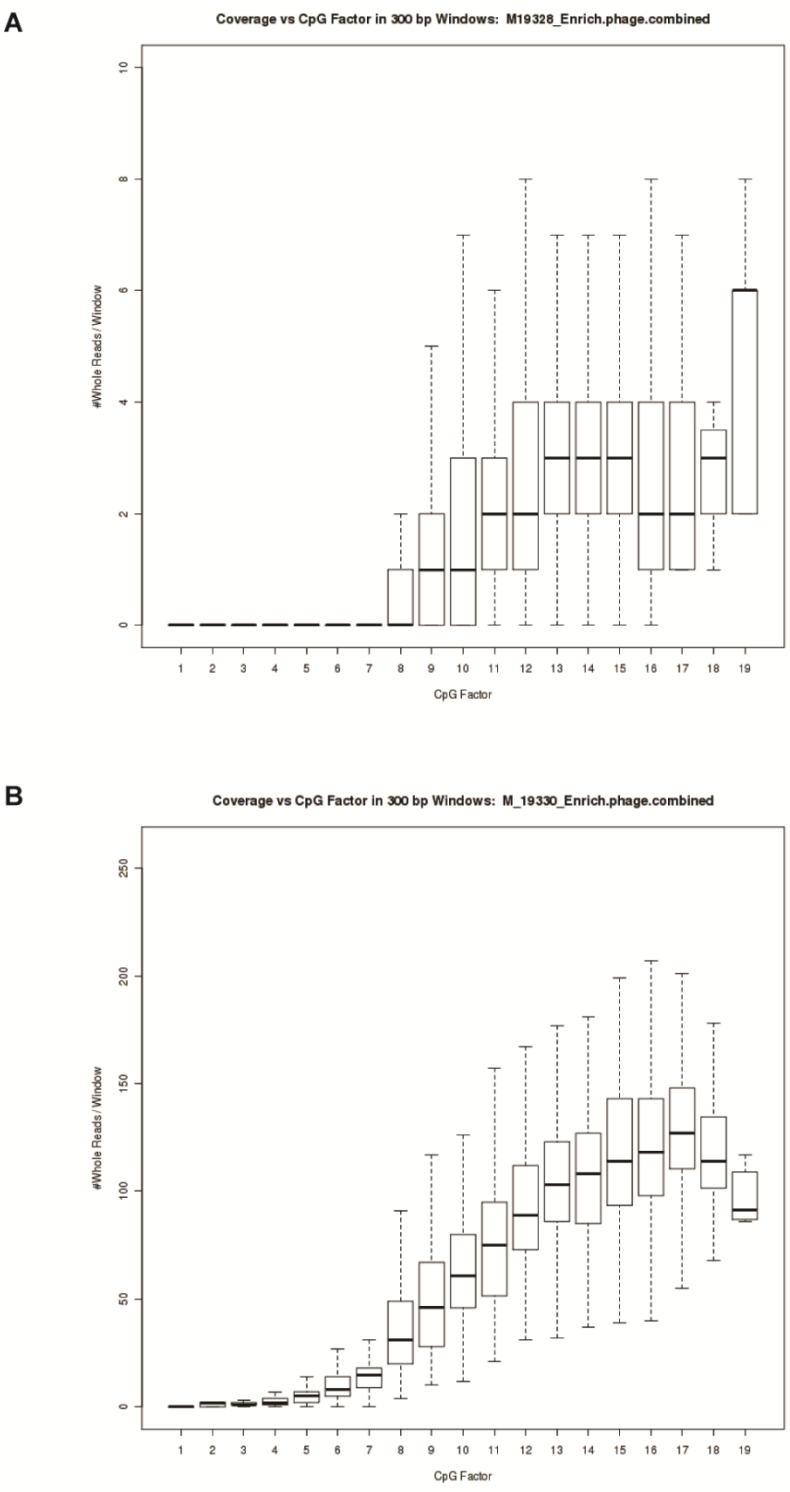


Figure 5.2: Representative methylated lambda phage sequencing results

Number of reads that align to lambda phage are plotted relative to their CpG density. **(A)** Example of a sample which does not appear to have sufficient reads for use as a standard control. **(B)** Example of desired lambda phage control results.

provided by the core facility suggesting that our workflow was working as intended (Figure 5.3).

We opted to move forward with the Minfi analysis beta values as they incorporated more rigorous normalization and our control gene results more closely matched expected trends. Quality assessment plots suggest our sample results are of high quality, following expected trends in the distribution of methylation/beta values (Figure 5.4). A cursory look at gene promoters previously established as methylated in prostate cancer (GSTP1, PTGS2 and APC (not shown)), confirms high beta values for all probes in these genes in the cancer samples (Figure 5.5). Furthermore, genes with CpG islands and high expression, such as GAPDH, displayed lower beta values.

We next wanted to confirm that all samples were behaving as expected. Principle components analysis (PCA) and multidimensional scaling plots (MDS), measuring the distance between sample methylation statuses by observing total variability between samples, suggested that tumor normal pair 6058 may have been switched (Figure 5.6). The Normal for this sample clusters with low grade tumors and the tumor clusters with normal samples. To ensure our analysis incorporated only high quality data we opted to remove this sample from all analyses. Another sample with abnormal results by PCA was 7155. While the sample did not follow expected clustering, we could not clearly label its tumor

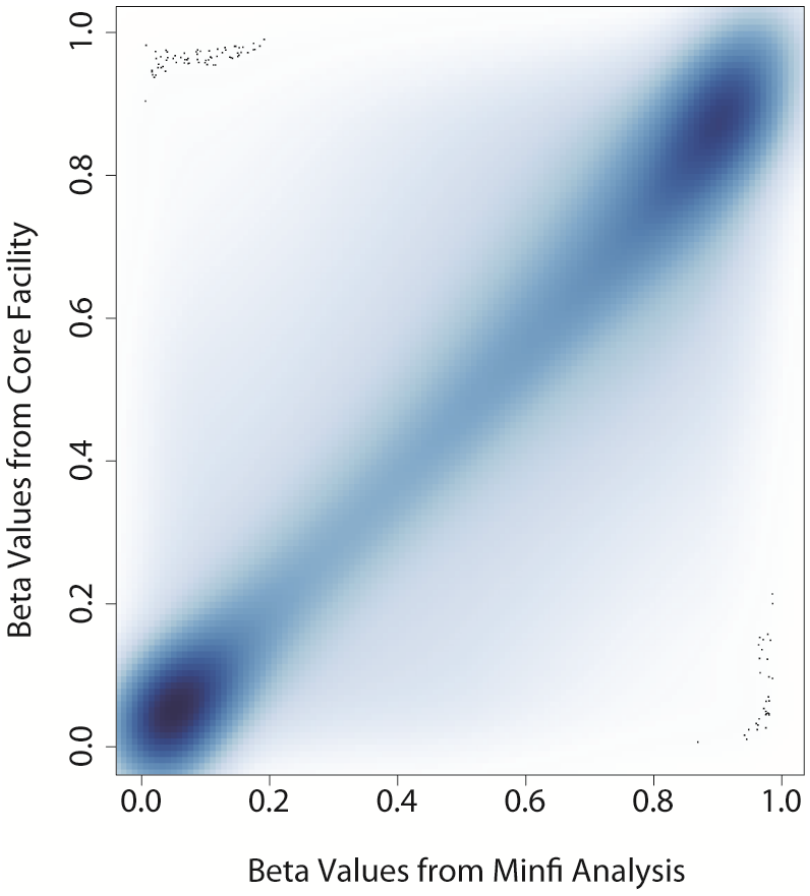


Figure 5.3: Comparison of core facility and Minfi package beta values

Beta values were calculated from the same data using two different methods.

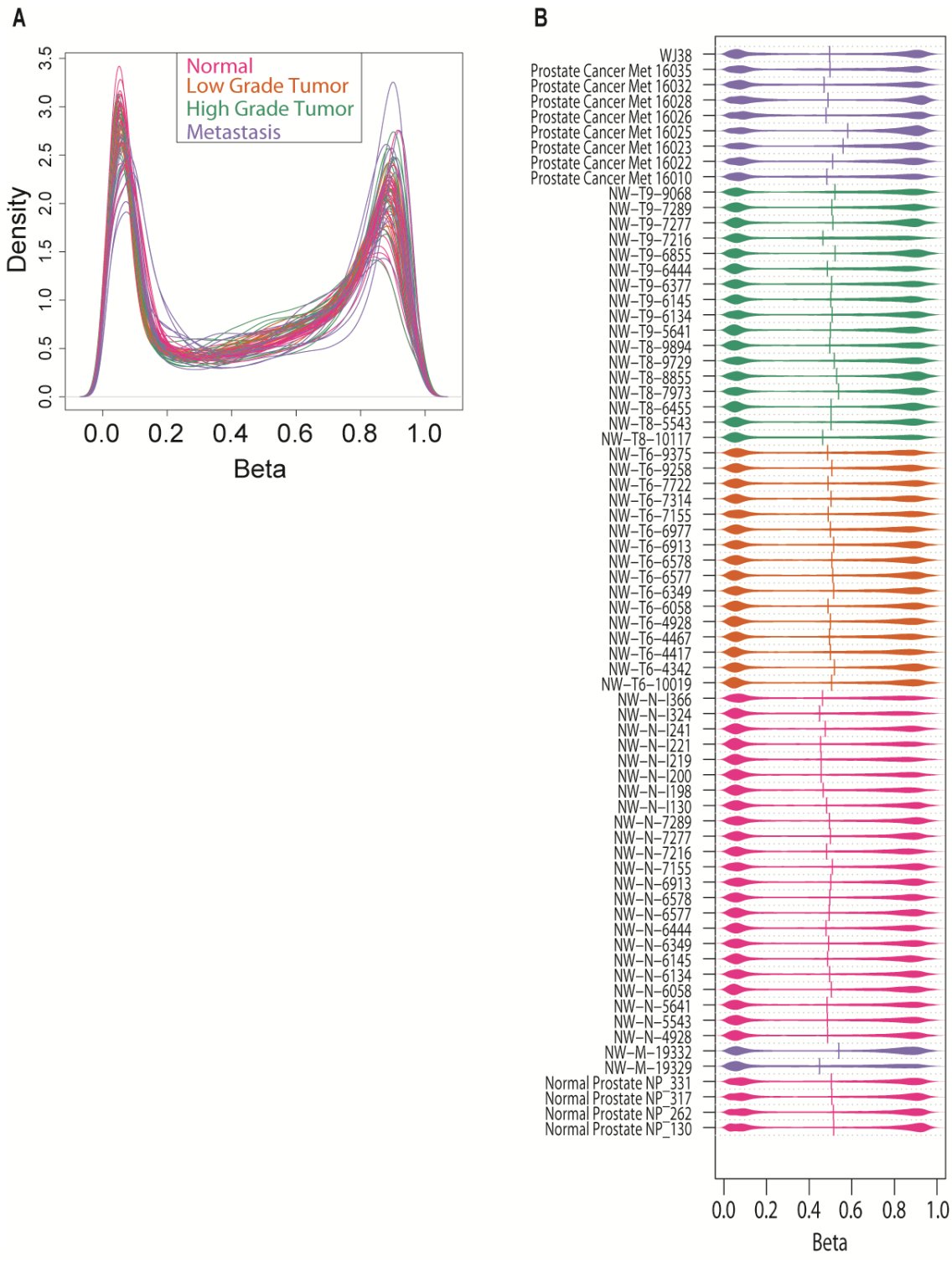


Figure 5.4: Quality control plots for Minfi package beta values

(A) Distribution of beta values for all samples. Note that the majority of values follow previously reported distributions and are very low (<0.2) or very high (>0.8). **(B)** Bean plot of each sample's beta value.

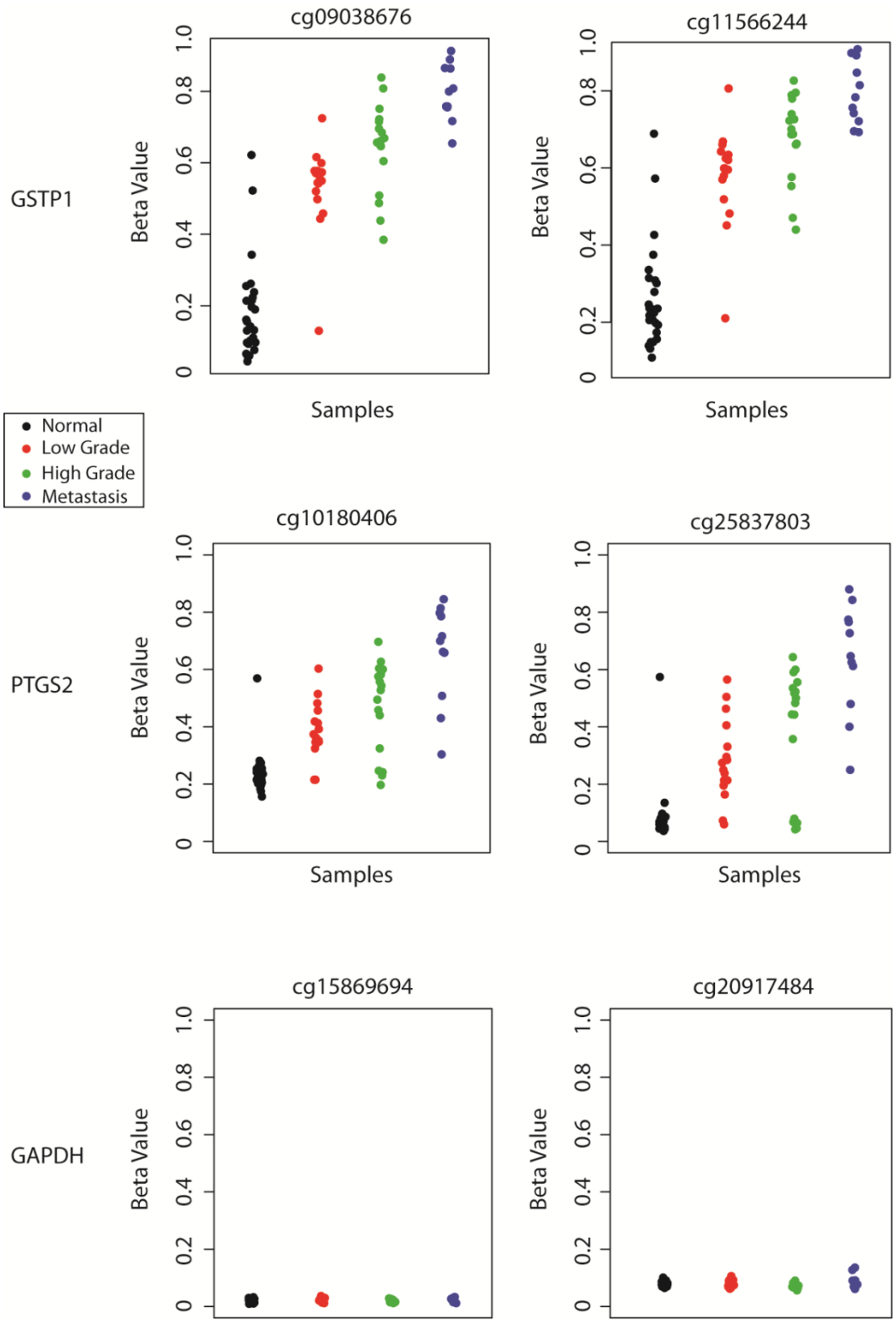


Figure 5.5: Beta values for representative probes of three control genes

GSTP1 and PTGS2 are known to have methylated promoters in prostate cancer and show increased methylation (beta value) with disease progression. GAPDH is unmethylated in all samples, as expected.

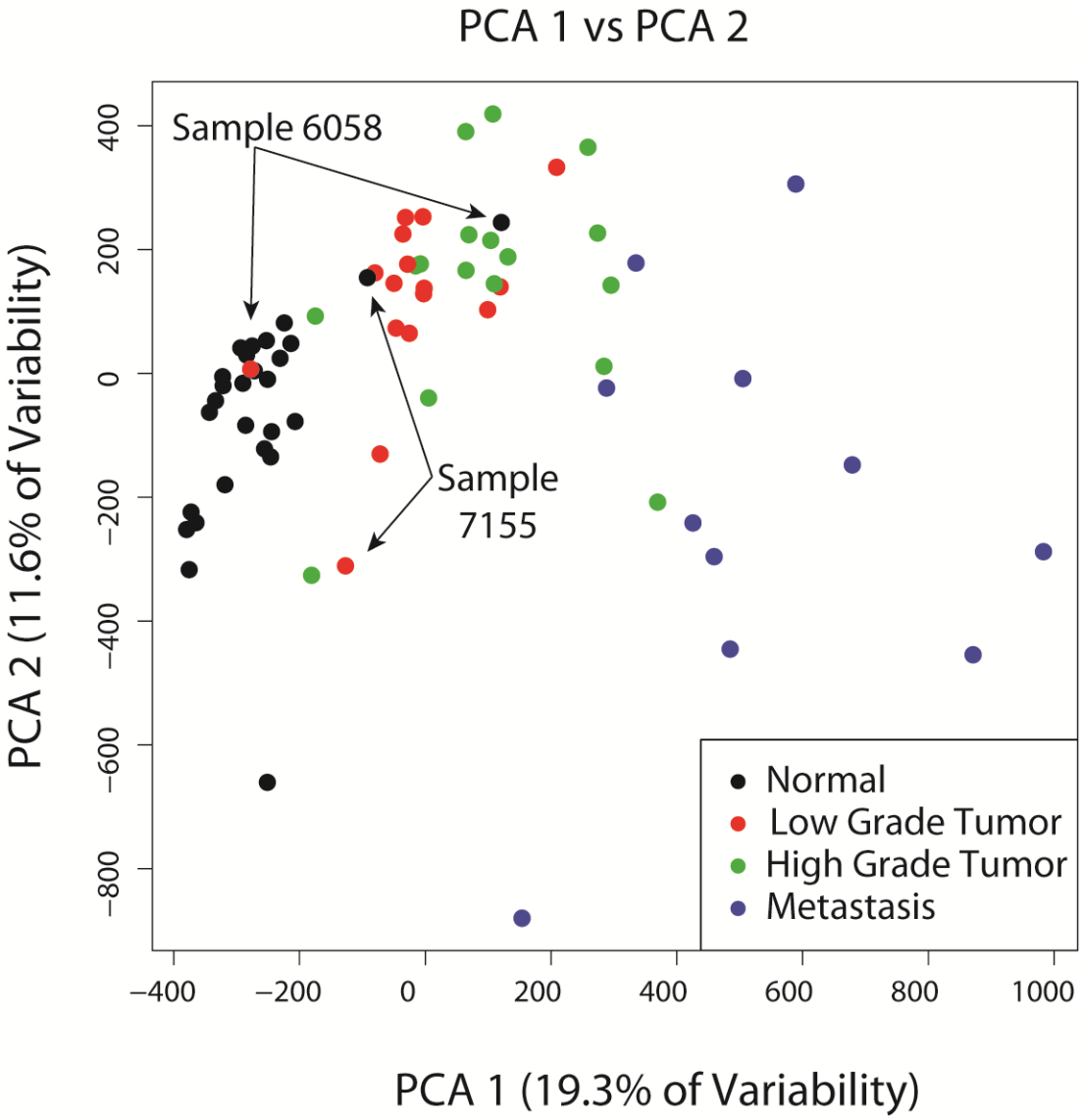


Figure 5.6: Principle components analysis with all 450K microarray probes

Sample 6058 and 7155 are highlighted to show abnormal clustering results. 6058 was removed from further analysis with this data set.

normal as switched, like 6058, and we decided to include it in our analysis.

Comparison of MBD-Seq and 450K Microarray Data

A key question entering this project was how the next generation sequencing based methylome analysis would compare with an established method, in this experiment the 450K Methylation Microarray. To answer this question we took the peaks identified by MACs calls in our sequencing data that overlapped (>1bp) with regions that contain a probe for our 450K methylation microarray and placed them in one of three bins: no peak (or p-value < 10^{-5}), low confidence (10^{-5} < p-value < 10^{-10}), or high confidence (p-value > 10^{-10}). To evaluate between sample groups the bins were formed individually for normal, low grade, and high grade samples. The 2 metastatic samples analyzed on both the microarray and sequencing platforms were incorporated into the high grade group. Comparison of the sequencing p-value bins and microarray beta values showed significant correlation (Figure 5.7). A histogram (Figure 5.8) confirmed that only a limited number of probes do not correlate well with sequencing. While the value of the Pearson correlation comparing p-values to fraction methylated is not relevant, it's significance supports our hypothesis that the regions identified by microarray as methylated are also found by MBD-Seq (Table 5.2).

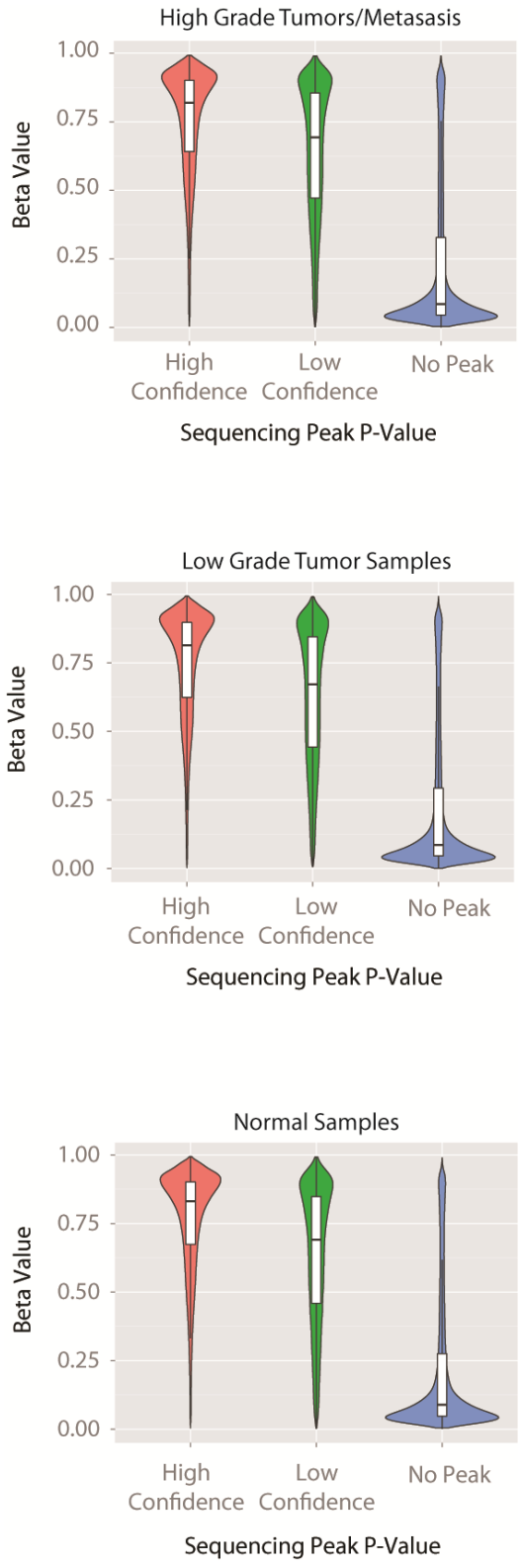


Figure 5.7: Violin plot of sequencing peak p-values and beta values from microarray

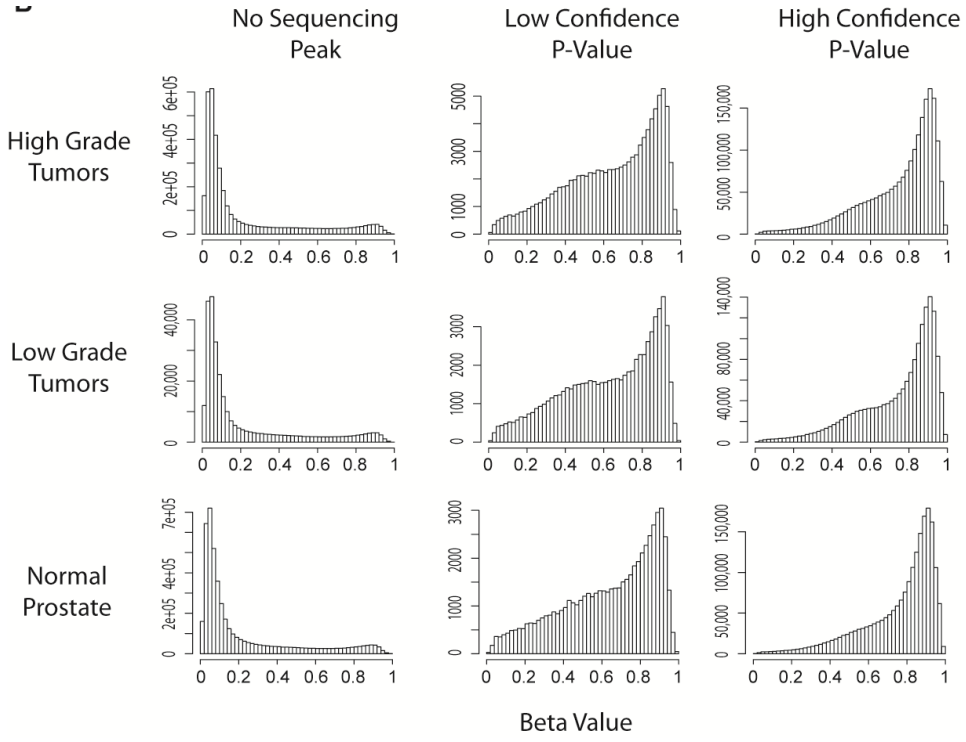


Figure 5.8: Comparison of methyl-sequencing peak p-values and microarray data

histogram showing correlation of sequencing peaks with microarray beta values for the same samples. Note for (A) surface plots in each column are the same, plotted on a different scale.

Sample Type	Pearson Coefficient	95% CI	P-Value
Normal Prostate	0.570	0.570-0.571	<2.2e-16
Low Grade Tumor	0.559	0.558-0.559	<2.2e-16
High Grade Tumor/Metastasis	0.536	0.536-0.537	<2.2e-16

Table 5.2: Pearson coefficients comparing sequencing peak confidences and microarray beta values

Initial Analysis of the 450K Methylation Microarray Data Across All Samples.

Having established that our microarray data was both high quality and significantly correlated to our sequencing data we set out to determine whether a panel of methylation marks which could assist with risk stratification in a Gleason dependent or independent manner. After removing one sample pair which was likely inverted (6058 normal and tumor) we removed all probes with a CpG density within a 300bp window around the probe location of less than 2.5% (Figure 5.9A). This was to avoid regions where the methylation status is likely influenced by a very limited number of CpG base pairs. This left us with approximately 311,000 probes. In a sample independent manner we sorted the probes based on their standard deviation (Figure 5.9B). We then isolated the top 1000 or 5% (about 20,000 probes) most variable probes, based on standard deviation, for our analysis. We hypothesized these regions would provide the best signature for differentiating tissue types.

We began by comparing the 5% most variable probes by principle components analysis (Figure 5.10A). Interestingly, considering just the most variable component, PC1, nicely separated the majority of high grade and metastatic samples from low grades and metastasis (Figure 5.10A). We performed hierarchical clustering with these probes

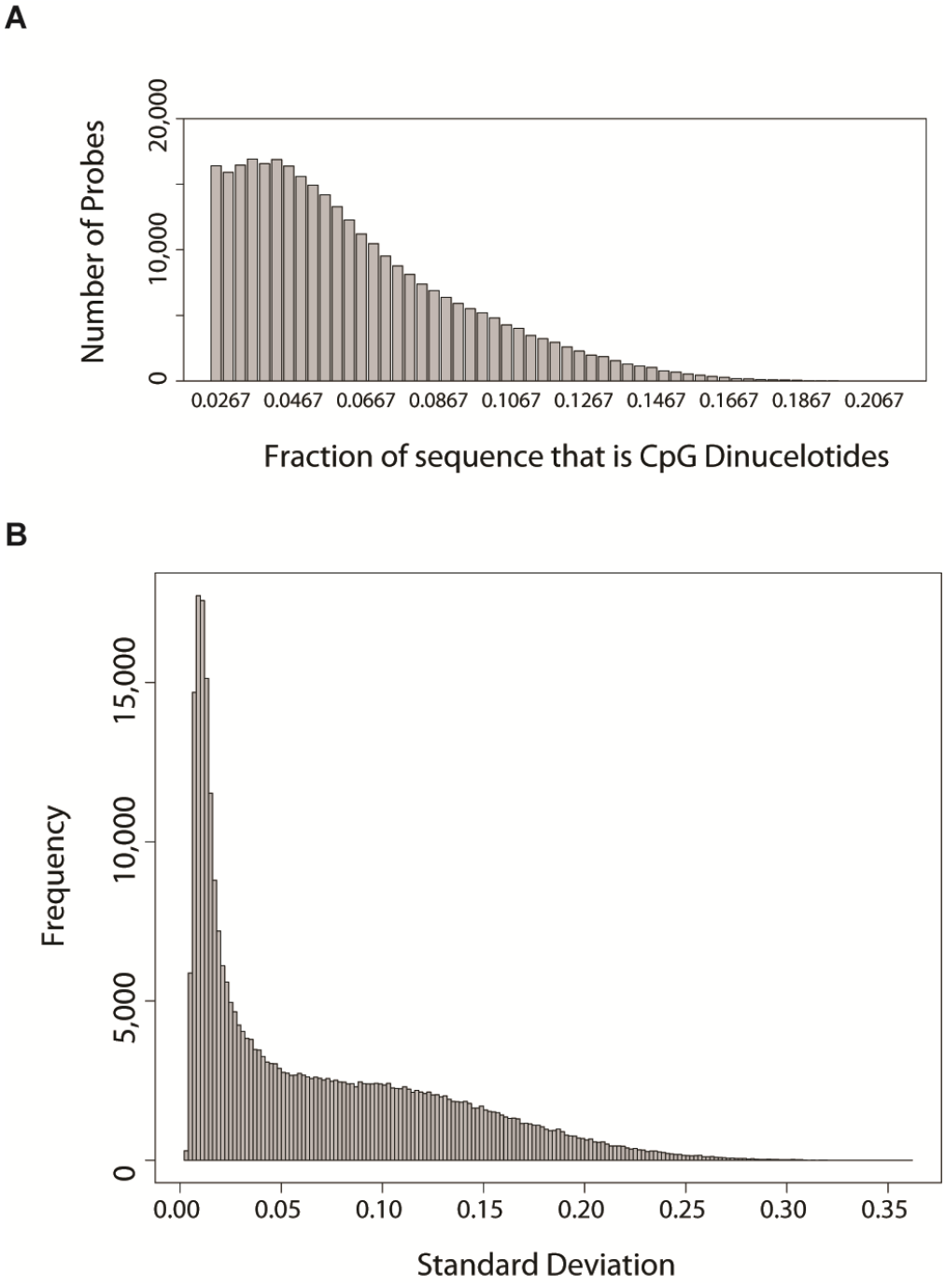


Figure 5.9: CpG density and standard deviation of probes used for finding biomarker panel

(A) Histogram displaying the fraction of total sequence that is a CpG dinucleotide after removing sequences with density <0.025. **(B)** Histogram showing standard deviation for all probes with CpG density>0.025 across all probes, independent of sample.

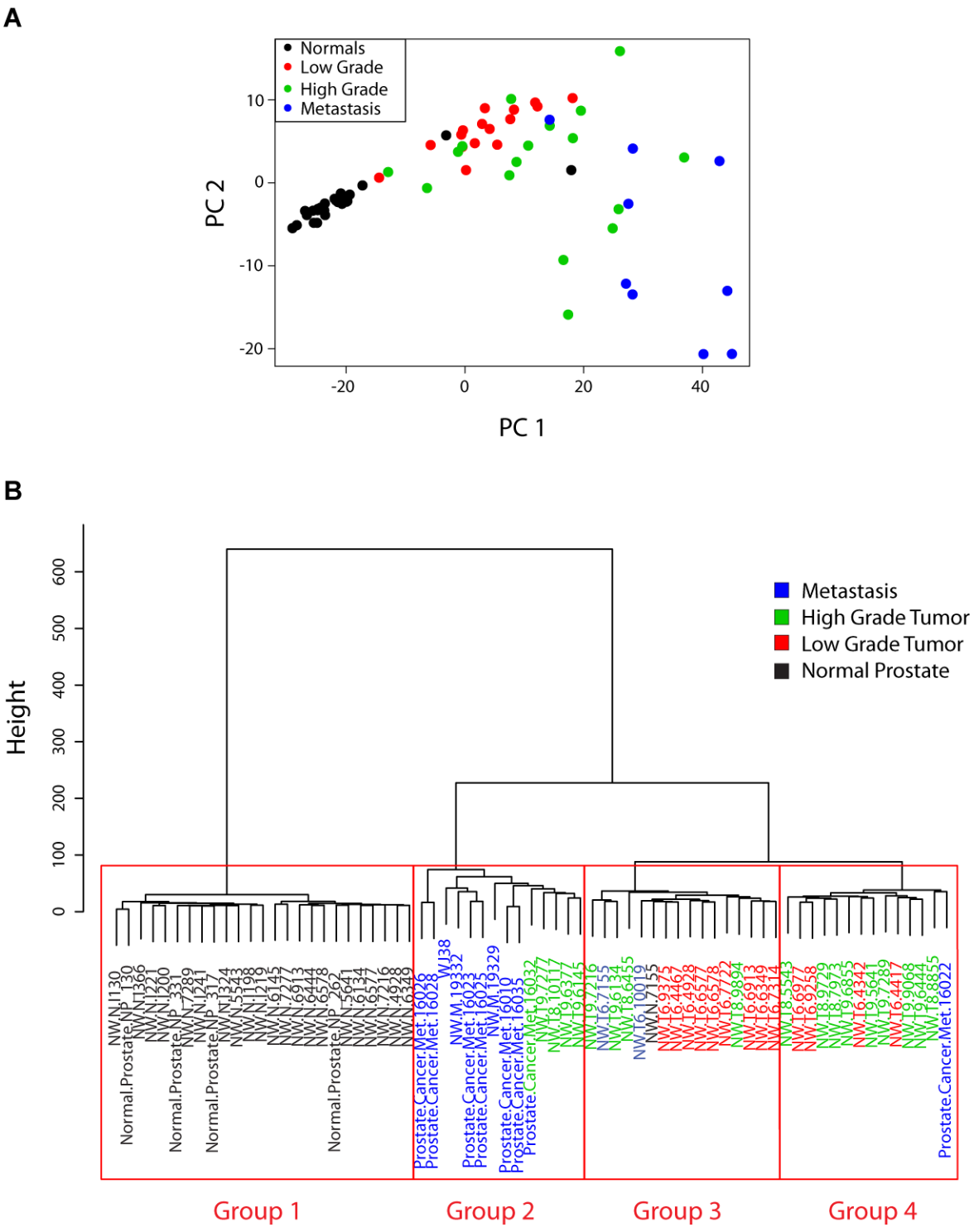


Figure 5.10: Initial evaluation of the 5% most variable microarray probes

(A) Principle components analysis and **(B)** hierarchical clustering of 5% most variable probes based on beta value standard deviation across all samples. Hierarchical clustering yielded 4 groups, with group 1 comprised only of normal tissue, and group 2 comprised of metastatic and high grade tumors, all likely to have a poor outcome.

and found they created four clusters of samples. The first is enriched for normal samples and clusters the farthest from the other three groups. This is unsurprising as it would be expected for normal tissue to contain similar methylation patterns, likely organized to maximize the function of prostate tissue. Group two was the second most unique identified, and contained exclusively metastatic and high grade tumors. This was extremely exciting as the outcome of patients with these tumors is often poor. We also observed that all metastasis with multiple samples from the same patient clustered together. A previous report using similar samples had the same finding of clonal maintenance across metastases, supplying some confidence that our analysis is sound(*135*).

Before attempting to identify a biomarker panel we wanted to query whether smaller groups of probes could distinguish between our Gleason scores. For this question we created heat maps of the top 1000 (Figure 5.11) and top 300 (Figure 5.12) most variable probes by standard deviation. Similar to the previous results with 20,000 probes, these maps clearly organized samples into a group containing predominantly normal prostate tissue, and another containing metastatic and high grade samples. A number of tumors now appear with the normal group, though, and closer analysis revealed that these tumors cluster with their respective matched normal tissue. Despite attempting to control for tumor content during pathology review, we suspect these cancers have significant normal stromal contaminant present in their frozen tissue

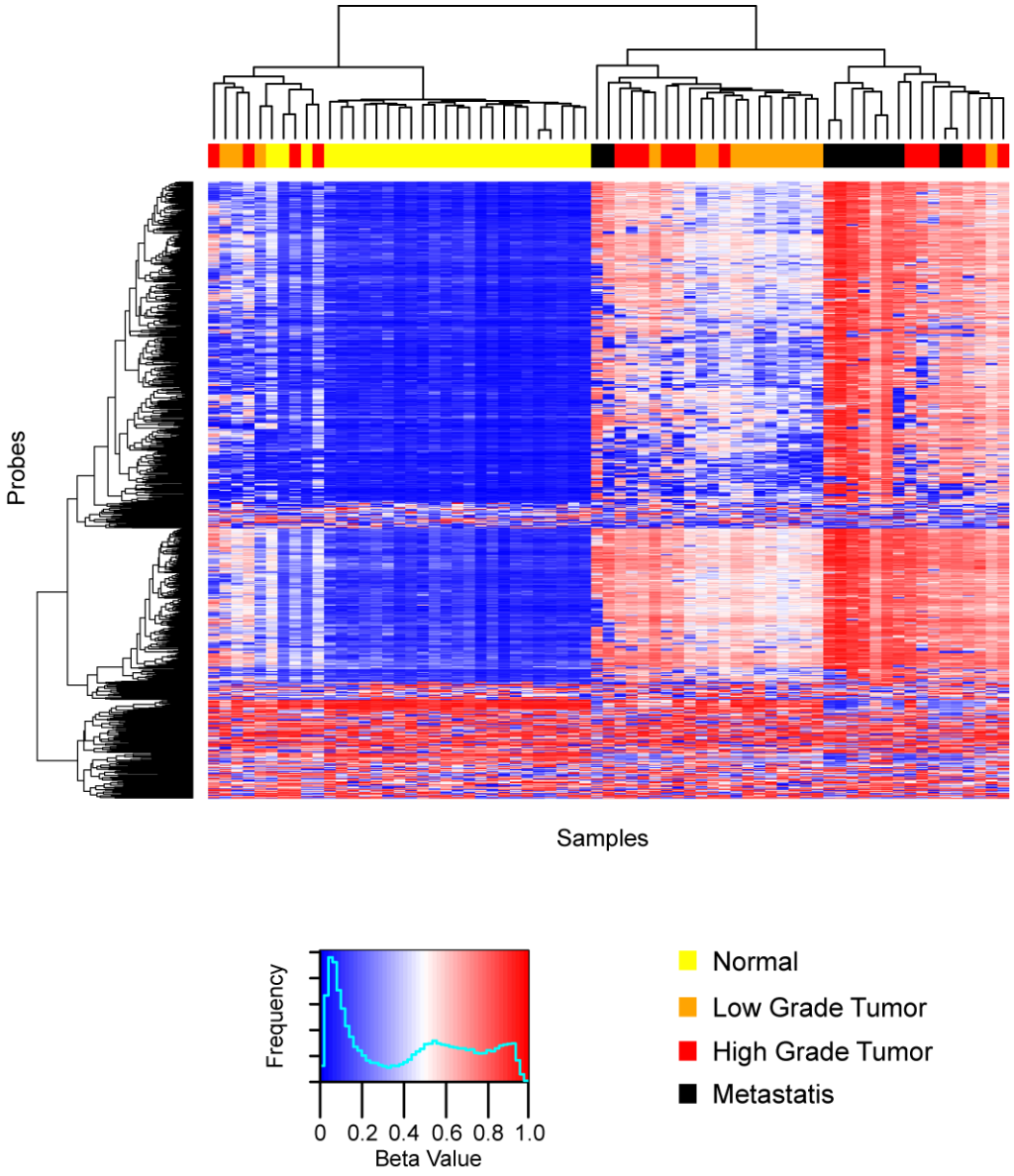


Figure 5.11: Heat map of the 1000 most variable microarray probes

Most variable probes based on standard deviation, independent of sample type.

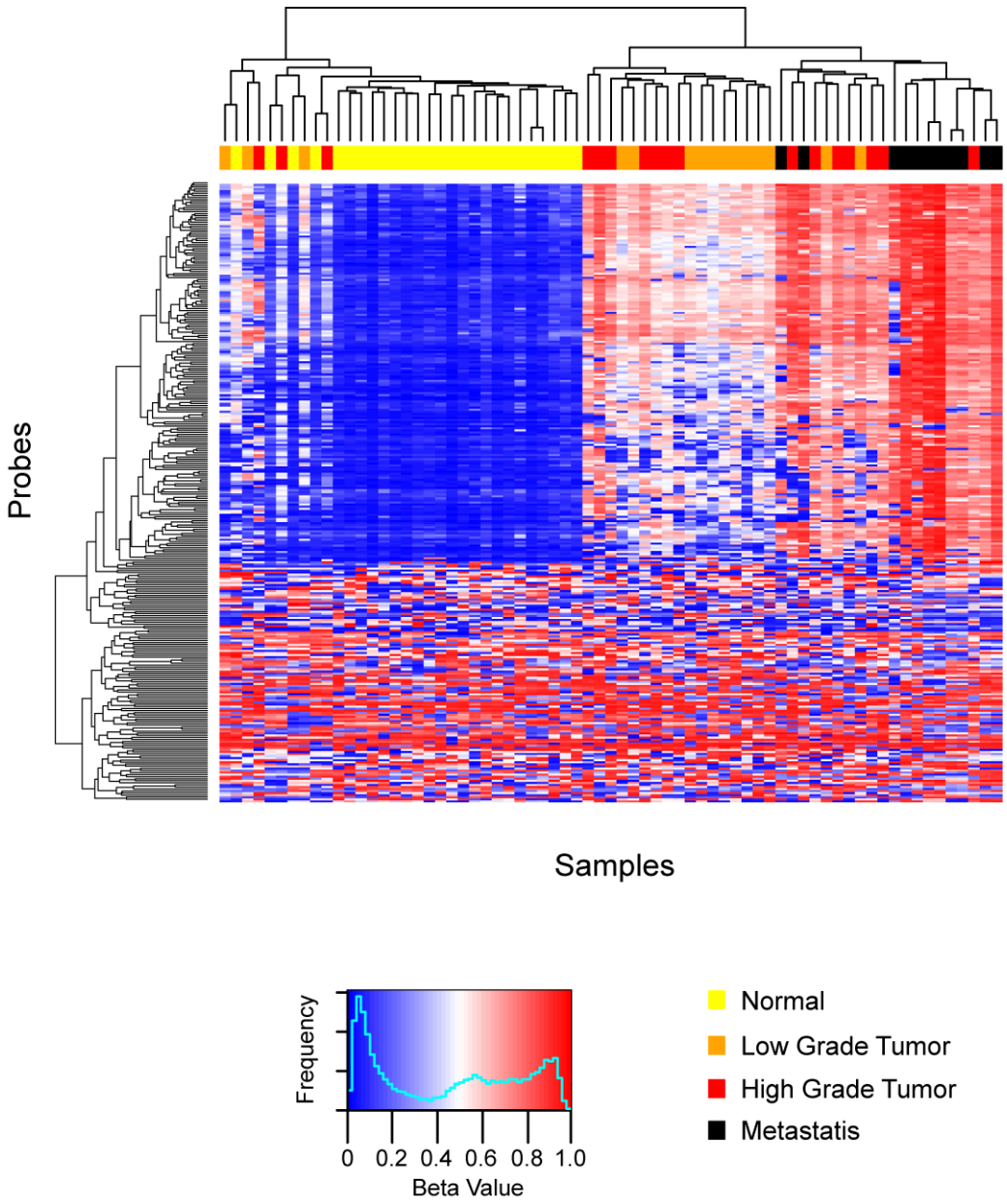


Figure 5.12: Heat map of the 300 most variable microarray probes

Most variable probes based on standard deviation, independent of sample type

leading to this mal-clustering. Another possibility is that these tumors have relatively few DNA methylation alterations compared to their matched normal tissue. Another interesting, but expected, observation is that the majority of probes selected by our analysis (which is largely restricted to regions with high CpG density) appear to gain methylation (increase their beta value). Previous reports have established that hypermethylation at specific gene promoters is better maintained than hypomethylation(*135*).

Identification of a DNA Methylation Biomarker Panel to Assist with Gleason Scoring

Confident that our analysis method could identify probes that differentiate our samples based on Gleason Score we attempted to identify a small panel of markers which could serve as a biomarker for Gleason grade. We first approached our question by using simple ANOVA comparisons between each probe and all four sample groups. Unfortunately this yielded probes which had very stable beta values, but not necessarily significantly different in magnitude between groups. To avoid this we returned to our 5% most variable probes and queried them in this simple manner: we calculated sensitivities for each probe, looking at high grade + metastasis vs. low grade tumors and vice versa. We eliminated all probes with sensitivities <0.25. We then calculated a

specificity for each probe, and excluded any which was <0.9. Interestingly, there were no probes that showed higher beta values for low grade samples with the sensitivity and specificity cutoffs we employed. Comparing high grade + metastasis to low grade, however, yielded 72 probes from 53 unique genes (Table 5.3). A boxplot of beta values for all 72 probes based on their sample type shows a significant relationship between increasing beta value and Gleason score/tissue stage (Figure 5.13A). Principle components analysis on the probes identifying these differentially methylated regions (DMRs) clearly distinguish metastatic and high grade samples from normal and low grades (Figure 5.13B). The one normal tissue outlier, indicated by an arrow, belongs to sample 7155 which was flagged earlier as possibly switched with its tumor (Figure 5.13B).

We next constructed a heat map to further investigate the properties of this panel. The heat map separated samples into 2 groups (Figure 5.14). The first group is comprised of all metastatic and high grade samples, similar to group 2 in our previous heat maps with the 1,000 and 20,000 most variable probes. The second group for this panel contains the remaining samples but accurately sub clusters them based on Gleason score with only a couple exceptions (Figure 5.14). To obtain a better idea of the biological functions of these genes we performed GSEA analysis and found that there was no significant enrichment for any pathways or families (data not shown). Analysis of the probes themselves

show that the majority locate to CpG islands, with only 14% found in gene bodies or intergenic regions (Figure 5.15 A and B). This data is supported by the short distances to transcription start sites for each of these probes (Figure 5.15C). Since the 450K methylation microarray is enriched for probes at promoter regions these results were not unexpected. We applied both the Tukey Honest Significant Difference and analysis of variance (ANOVA) statistics to confirm the significance of our panel, observing p-values of <2.2e16 for the differences in methylation status (beta values) displayed by the panel (Figure 5.16).

Gene	# of Probes	Sensitivity	Specificity	Entrez ID	Probe	Probe Relation To Promoter
ADHFE1	6	0.29	0.93	NM_144650	Multiple	Island/S_Shore
BHLHB9	3	0.43	0.93	NM_001142528	Many	N/A
C20orf103	3	0.25	0.93	NM_020341	Many	Island
COL11A2	3	0.32	0.93	NM_021976	Many	N_Shore
TTTY14	3	0.43	0.93	NM_001173413	Many	Island
CYBRD1	2	0.32	0.93	NM_024843	Many	N_Shore
HIST1H2BE	2	0.25	0.93	NM_003523	Many	N_Shore
PLEC	2	0.25	0.93	NM_201384	Many	N_Shelf
PTGDR	2	0.29	0.93	NM_000956	Many	Island
TINAGL1	2	0.32	0.93	NM_001525	Many	N/A
ZSWIM2	2	0.25	0.93	NM_005795	Many	Island
ADAMTS1	1	0.36	0.93	NM_007038	cg15621322	Island
AEBP1	1	0.25	1.00	NM_006230	cg14249876	N/A
AIFM2	1	0.25	1.00	NM_173555	cg17446739	Island
ASCL2	1	0.39	0.93	NM_005170	cg13930892	Island
AXIN2	1	0.29	0.93	NM_004655	cg17797591	N_Shore
BCAT2	1	0.25	0.93	NM_016246	cg18873530	Island
C19orf77	1	0.25	1.00	NM_001136503	cg08380311	Island
C6orf141	1	0.29	0.93	NM_000324	cg26081162	Island
C6orf150	1	0.25	1.00	NM_138441	cg00463577	Island
CCL28	1	0.25	0.93	NM_148672	cg04187185	Island
DDIT4L	1	0.25	0.93	NM_145244	cg27062369	Island
DSC3	1	0.25	0.93	NM_004949	cg03017520	Island
EFCAB4B	1	0.25	0.93	NM_032680	cg17904739	N_Shore
EPM2AIP1	1	0.29	0.93	NM_014805	cg05670953	N_Shore
FBXO47	1	0.29	0.93	NM_020405	cg14399930	N/A
FOXE3	1	0.25	0.93	NR_026878	cg19809499	Island
GNG4	1	0.29	0.93	NM_001098721	cg17816394	Island
H1FNT	1	0.32	0.93	NM_181788	cg00201175	N/A
HEXA	1	0.29	0.93	NM_001080462	cg11540416	S_Shore
HLA-A	1	0.29	0.93	NR_028032	cg11808100	Island
LAMC2	1	0.25	1.00	NM_005562	cg01949993	N/A
LHX3	1	0.29	1.00	NM_178138	cg12250872	Island
MSN	1	0.29	0.93	NR_029637	cg24923543	Island
NFIL3	1	0.29	1.00	NM_005384	cg07084627	Island
NKAP	1	0.25	0.93	NM_032498	cg25908030	Island
NR0B1	1	0.25	1.00	NM_000475	cg11869614	Island
PHOX2A	1	0.25	0.93	NM_030813	cg01670677	Island
PRKAR1B	1	0.32	0.93	NM_002735	cg12441126	Island
PTGS2	1	0.25	0.93	NM_024420	cg23070111	Island
PYCARD	1	0.25	0.93	NM_145182	cg05361373	Island
RASL10A	1	0.32	0.93	NR_003686	cg08105965	Island
RND1	1	0.29	0.93	NM_033124	cg27241190	N/A
SFRP2	1	0.25	0.93	NM_017639	cg14330641	Island
SLC40A1	1	0.25	1.00	NM_014585	cg10752008	N_Shore
TBX3	1	0.25	1.00	NR_030351	cg11412853	N_Shore
TICAM2	1	0.32	0.93	NM_181836	cg08870042	Island
TRIM46	1	0.36	0.93	NM_001044390	cg15497761	Island
TSPYL3	1	0.25	1.00	NR_002781	cg07071389	Island
USP44	1	0.36	0.93	NM_032147	cg13879483	Island
YPEL3	1	0.36	0.93	NM_031477	cg27106909	N_Shore
ZDHHC1	1	0.25	0.93	NM_013304	cg00420361	Island
ZNF311	1	0.25	0.93	NM_030903	cg25255293	Island

Table 5.3: Genes with at least 1 probe in our Gleason score biomarker panel

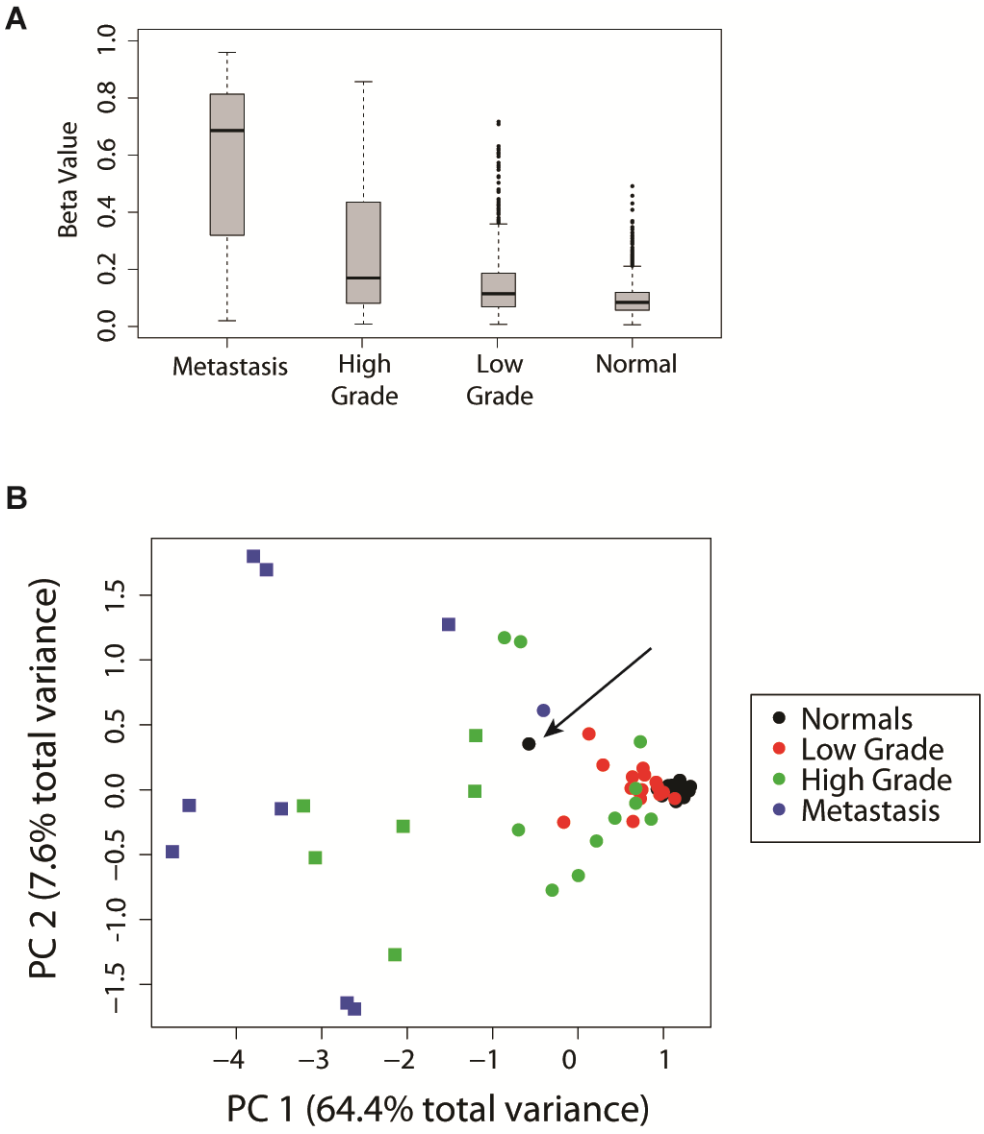


Figure 5.13: Initial Analysis of Gleason score differentiating DNA methylation biomarker panel

Initial Analysis of novel DNA methylation biomarker panel distinguishing prostate cancer samples based on Gleason score and stage. **(A)** Boxplot describing beta values based on samples and **(B)** principle components analysis across all 72 probes that identify the differentially methylated regions of this panel. The arrow in (B) indicates sample matched normal 7155 which may be switched with its tumor sample.

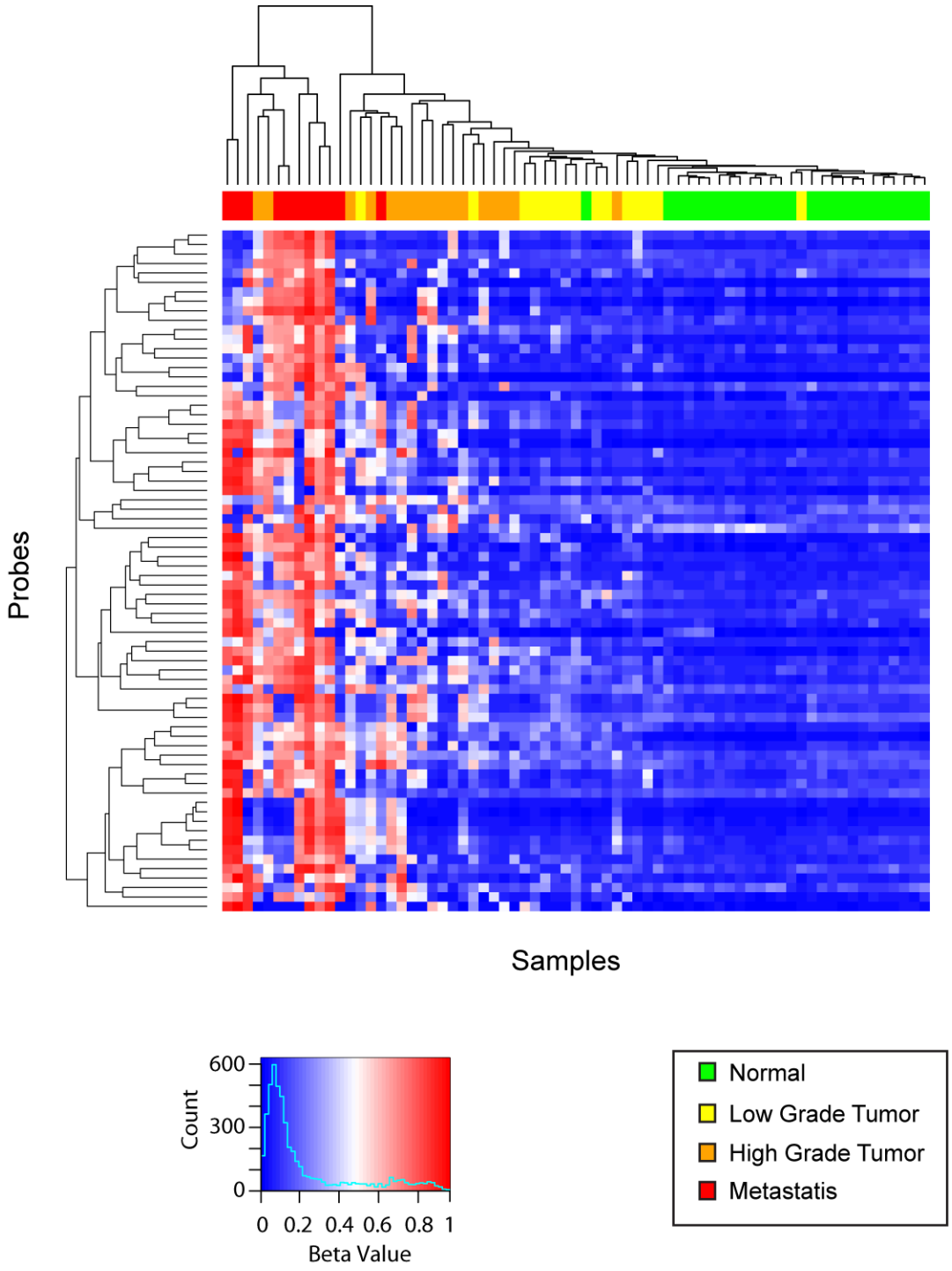


Figure 5.14: Heatmap of 72 probes comprising the Gleason score DNA methylation biomarker panel

Heatmap of all probes in our DNA methylation biomarker panel for distinguishing prostate cancer samples based on Gleason score and stage.

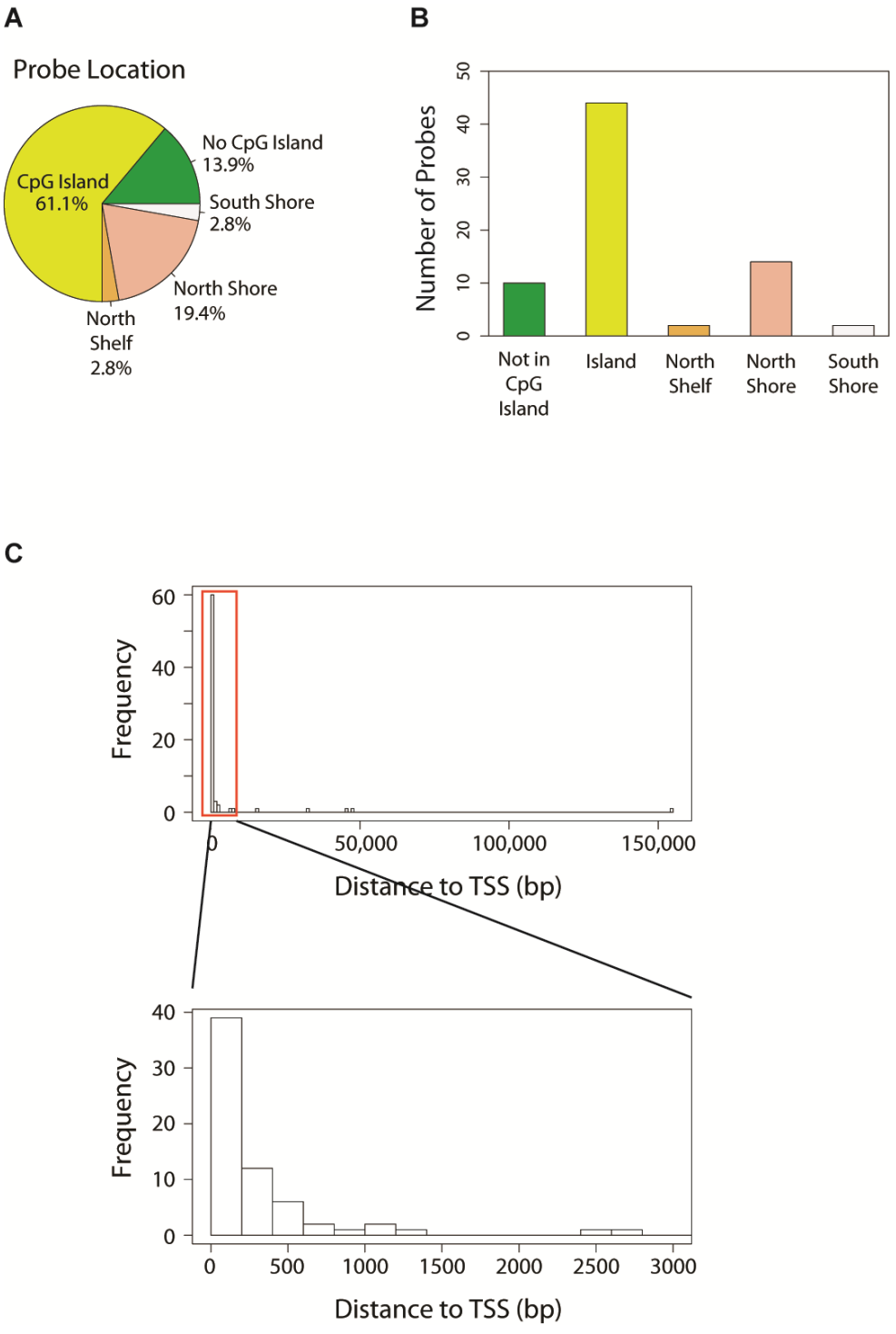
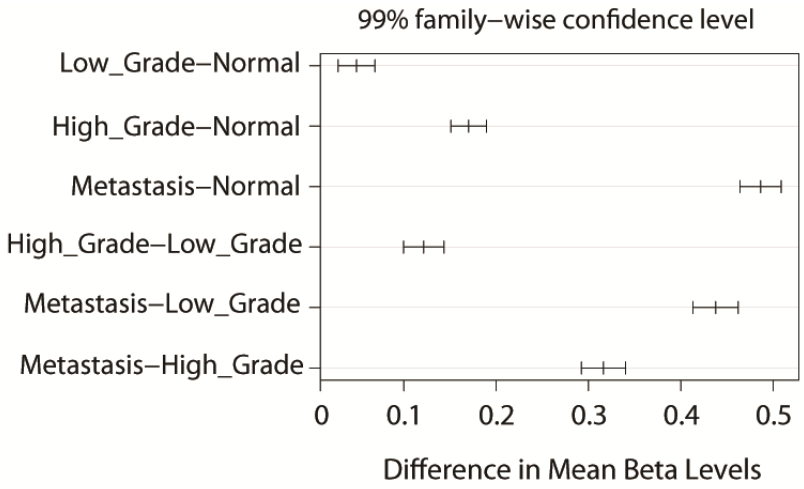


Figure 5.15: Summary of probe locations in Gleason score DNA methylation biomarker panel

Summary of novel Gleason score differentiating methylation biomarker panel probe locations. **(A)** Pie chart and **(B)** bar graph describing probe locations. **(C)** Histogram showing distance to transcription start site. The majority of probes are in CpG islands, illustrated by the small distances shown.



Comparison	Delta	Lower 95% CI	Upper 95% CI	P-Value
Metastasis-Normal	0.486	0.468	0.505	<2.2e-16
Metastasis-Low_Grade	0.438	0.417	0.458	<2.2e-16
Metastasis-High_Grade	0.316	0.297	0.336	<2.2e-16
High_Grade-Normal	0.170	0.154	0.186	<2.2e-16
High_Grade-Low_Grade	0.121	0.103	0.140	<2.2e-16
Low_Grade-Normal	0.049	0.032	0.065	<2.2e-16

Figure 5.16: Statistical tests comparing sample groups based on Gleason score biomarker panel

Summary of statistical tests confirming significance of beta value differences observed in our novel biomarker panel distinguishing prostate cancer samples by Gleason score and stage.

5.4 Discussion

Since the discovery of PSA in 1979 by Wang et al, and its entry into the clinic in the mid 1990's, the field of urology has benefited from one of the best biomarkers ever discovered for cancer diagnosis (*183*). With thorough pathology and staging, accurate prognosis often follows diagnosis (*184*). Unfortunately pathology is always limited by how well the biopsy actually samples the tumor. Since no good radiological methods for guiding biopsies currently exist, the only direction the needle is receiving is luck. To investigate a solution to this problem we analyzed DNA methylation levels in prostate cancer and normal tissues to show that they could serve as an accurate biomarker to assist with Gleason scoring.

Application of a novel technology requires a thoughtful and meticulous approach to avoid missing biases or misleading results. Here we carefully compare methylation levels derived from a microarray and sequencing platform from 69 samples to show that our MBD-Seq approach is both accurate and reproducible. Highly significant correlations between the data, as measured by Pearson coefficients of >0.5, provide statistical evidence that our methods are highly correlated and thus serve to independently cross-validate each other. The MBD-Seq method is especially exciting because of its unbiased approach. Unlike microarrays, which are limited by probe locations that are enriched at CpG islands and promoters, and bisulfite sequencing, hindered by

conversion rate efficiencies and high cost, MBD-Seq provides a bias free approach to analyzing the entire methylome. Bioinformatic efforts are ongoing to identify differentially methylated regions with these data.

In the meanwhile, we have gone forward with using the Infinium 450K microarray data. We initiated this work by removing almost 25% of the probes in the array because of low CpG densities. These regions contained less than 8 CpG's within a 300bp window of the CpG that was queried. These regions concerned us because they were not representative of regions where CpG methylation is likely to have a significant impact on transcript levels. Removing them helped us enrich our discovery set for regions which likely had biological consequences as a result of their methylations status.

Based on the problems observed with simple ANOVA comparisons we used the opposite methodology, and only considered probes with the high variability across all samples. After applying some stringent cutoffs to specificity we identified a small, 72 probe, panel that could distinguish between normal prostate, low grade, high grade, and metastatic samples. Despite comparing high grade and metastasis to low grade and vice versa, we found only probes with methylation that increased with grade/stage were significant. This is also apparent in our heat maps of the 72, 300 and 1000 most variable probes (Figures 5.14, 5.11 and 5.12). This finding is supported by recent evidence that hypermethylated

regions are better conserved in disease progression than hypomethylated ones (*135, 136, 177*).

Another interesting observation is detected when looking at the confidence intervals and statistical testing (Figure 5.16). Despite designing this experiment to find biomarkers that distinguish between Gleason scores/stage we achieved a panel which strongly separates metastatic samples compared to primary and normal tissue (Figure 5.14 and 5.16). It is unclear why this occurred, but it would be interesting to investigate the genes identified in the biomarker panel more closely to determine what biological benefit they may offer a metastatic cell. Such an analysis could yield a novel therapeutic target to prevent or slow transition to the metastatic state. Another tantalizing observation is found with the clustering of high grade tumors, which form two groups: one associated with metastasis and another with low grade tumors. It would be interesting to determine if the high grades that cluster closer to low grade and normal prostate tissue are the patients who end up with less disease burden and better outcome despite poor pathological prognosis. This clustering could also be explained by stromal contamination, and further work will be necessary to investigate both of these possibilities.

In summary, we show that DNA methylation could be used to assist with Gleason scoring and tumor staging and may in fact be an independent predictor of risk. We also demonstrate that our novel

sequencing method, MBD-Seq, accurately measures whole genome methylation levels when compared to current gold standard techniques, and offers a bias free view of the methylome. Future students will continue work with the MBD-Seq results so that it can be used as a discovery set to form a methylation based biomarker panel to help with prostate cancer risk stratification.

5.5 Methods

Sample Acquisition and Quality Control

All primary fresh frozen samples were from patients treated at Johns Hopkins Hospital undergoing radical prostatectomy procedure. Collection was done by The Prostate Cancer Biorepository Network (PCBN) using pre-established protocols to minimize experiment error due to storage or handling(*185*). Metastasis samples were obtained through the Project to Eliminate Lethal Prostate Cancer (PELICAN) rapid autopsy program as previously described(*186*). For all primary samples H&E was reviewed by a single expert pathologist to confirm Gleason Grade (either 6 or >7), percentage of tumor (>70%), absence of cribriform or other significant inflammatory markers, or the presence of tumor in normal samples. Once this was confirmed 50-100 20μm slices were taken and the last slice was also reviewed by the same pathologist to confirm that the pathology had not changed while removing material from the block. DNA and RNA were extracted from these slices using a DNeasy (Qiagen) or RNeasy (Qiagen) kits. Samples were quantitated by NanoDrop (Thermo Scientific) and quality was confirmed by assessment with real time PCR using primers for ribosomal component 18S and beta globin (BioRad). 5μg of material was provided for this study. The DNA was requantified by NanoDrop. Double stranded DNA concentration was determined using a fluorescence based Qubit® assay following manufacturer's instructions (Life Technologies). 100ng of each sample was run on a 2% agarose gel to

determine average size and quality. Samples that required re-purification due to possible contamination were run through a DNeasy kit (Qiagen) per manufacturer's instructions and eluted into 100μL of elution buffer.

Library Preparation and Sequencing

Library preparation for next generation sequencing with ABI 5400 sequencers (Applied Biosystems) was performed per manufacturer's instructions with the following changes(*187*). 1-2 μg of DNA was sonicated to a modal size of ~150-250bp, and end-repaired using the NEBNext SOLiD DNA library preparation kit end-repair module following the manufacturer's protocol (New England Biolabs). After column-purification (using the Qiagen PCR purification kit), SOLiD P1 and P2 adapters lacking 5' phosphate groups (Life Technologies) were ligated using the NEBNext adapter ligation module and column-purified, and subjected to isothermal nick-translation by treating with Platinum Taq polymerase to remove the nick. The resulting library was divided into two fractions, a total input fraction, and an enriched methylated fraction. The enriched methylated fraction was then subjected to affinity enrichment of methylated DNA fragments by using recombinant C-terminal 6xHis-tagged MBD2-MBD polypeptides immobilized on magnetic beads, similar to previously described methods (*177*). The resulting enriched methylated fraction and the total input fraction were then subjected to library amplification using the NEBNext amplification module according to the

manufacturer's protocols, using 4 - 6 cycles for the total input, and 10 - 12 cycles for the enriched methylated fractions. Library fragments that were between 200 - 300bp were size selected after agarose gel electrophoresis. The libraries were then subjected to emulsion PCR and bead enrichment following the SOLiD emulsion PCR protocol (Life Technologies). The resulting beads were then deposited on the SOLiD flow cell and subjected to massively parallel 50bp single-read sequencing on a SOLiD v4.0 or 5500XL sequencer to achieve between 10 – 100 million reads for each sample. Reads were aligned to the reference human genome (version hg18) using default settings in Bioscope v1.3 software, with the exception of the bam output method, which was changed to alignment score. Methylated regions were identified as peaks of aligned sequencing tags using MACS v1.4 software (*188-190*), which allows identification of peaks after accounting for both global and local biases using the total input fraction.

For the array 500ng of material was provided to the Sidney Kimmel Comprehensive Cancer Center (SKCCC) Microarray core facility. Samples were prepared and methylation determined using the Infinium HumanMethylation450 BeadChip Kit (Illumina). Raw fluorescence data was provided to us by the core facility and processed internally.

Analysis of MBD-Seq Data

This work is still ongoing and not completed at this time.

Processing and QC of Illumina 450K Methylation Array Data

Raw microarray data was preprocessed and beta values were obtained using the Minfi Bioconductor package with their recommended settings, including subset quantile normalization(*182*). All analysis was performed using the open source software R v3.0.2(*165*).

Comparison of MBD-Seq and 450K Microarray Data

Beta values provided by the core facility were compared to beta values obtained from our own analysis using R v3.0.2(*165*). Comparison between sequencing and microarray data was performed using the Bioconductor packages Granges (*191*), IRanges (*191*), ggplot2(*192*), and vioplot (*193*) in R v3.0.2(*165*). Pearson correlation coefficients were calculated using R package Hmisc (*194*).

Initial Analysis of the 450K Methylation Microarray Data Across All Samples.

Analysis of 450K microarray beta values was performed using R v3.0.2(*165*) and the following packages: ggplot2 (*192*), Hmisc (*194*), cluster (*195*), and gplots (*196*)

. Sample 6058, tumor and normal, were removed due to suspicion they may have been interchanged. Analysis was performed only on probes

with CpG densities >0.025 within a 300bp window surrounding the probe location.

Identification of a DNA Methylation Biomarker Panel with 450K Methylation Microarray Data

All analysis was performed in R v3.0.2(*165*) with packages Hmisc (*194*), cluster (*195*), and excel (Microsoft). Only probes with CpG densities >0.025 within a 300bp window surrounding the probe location were considered. Probes were ordered based on their standard deviation across all samples, and the top 300, 1000, or 5% (approximately 20,000) probes were analyzed. To determine a biomarker panel we calculated sensitivity and specificity for the top 5% most variable probes by comparing high grade and metastasis samples to low grade samples. The inverse comparison was also performed. We then eliminated all probes with a sensitivity<0.25 and specificity<0.9. The resulting panel was visualized using R v3.0.2 and significance determined using statistical tests Tukey's Honest Significance Difference test and analysis of variance (ANOVA) test on the beta values binned into groups (metastasis, high grade, low grade and normal prostate).

5.6 Acknowledgements

I would like to acknowledge the Next Generation Sequencing Core, in particular Jennifer Meyers, Lauren Ciotti, Sarah Wheelan, Vasan Yegnasubramanian, Anuj Gupta, Alyza Skaist and Micheal Rongione for their help with the MBD-Seq project. I would also like to thank the great Dr. Michael Haffner for troubleshooting and optimizing the MBD-Seq method. The Illumina 450K Methylation microarray experiment was performed by Dr. Wayne Yu from the SKCCC microarray core. Finally I want to thank my mentor Dr. Srinivasan Yegnasubramanian for the many hours of time he spent discussing and planning this experiment. Biomarker discovery is not an easy field to learn and much hand holding took place.

This work was funded in part by the NIH/NCI Prostate SPORE grant, the Prostate Cancer Foundation, the Commonwealth Foundation, and the Department of Defense Prostate Cancer Research Program Pre-doctoral training grant (W81XWH-11-1-0618).

Chapter 6

Concluding Remarks

Before 2000 chemotherapy, radiation and surgery constituted the majority of cancer therapy provided in the United States. In the mid 2000's there was a flurry of activity to find small molecule inhibitors of specific genomic mutations, such as BCR-Abl, EGFR and Her2. This was a consequence of the enormous efforts at Johns Hopkins, Harvard and other universities to obtain high coverage sequencing information on numerous cancer types. The flood of novel genetic information led to identification of numerous novel targets for pharmaceutical companies and biotech. Direct inhibition of these mutant proteins significantly abrogated many different diseases, offering a new treatment regimen for oncologists. Small molecule targeting was often simplified by the fact that the genomic alterations created disease specific pockets which normal tissue would not harbor(*197*). This avoided untoward side effects, allowing physicians to avoid dose limiting toxicities often seen with radiation or chemotherapy.

Every story has another side, though, and the bane of targeted therapy has and continues to be subsequent selection for a small molecule resistant tumor subtypes and regression (*197-199*). A cursory look through the literature shows that in the sea of people who have complete and maintained response to targeted therapy, there are oceans of patients who suffer recurrence (*197-200*). Efforts continue in this field to find novel, second generation small molecules, to overcome resistance or to combine with primary therapy to avoid resistance all together.

As an alternative path to genetic therapy, in this thesis I describe targeting of another class of alterations that drive oncogenesis, epigenetics. As discussed many times here a number of existing groups investigate this problem, with a handful of FDA approved successes. We hoped to create an advantage for ourselves by solving the MBD2 crystal structure and using it for structure aided design. While this never materialized, we did successfully obtain diffracting crystals and collected enough data to generate reasonable electron density maps. Given previous difficulties by ourselves and others, finding suitable conditions for crystallizing a methyl binding domain family member in isolation for structure aided design may prove difficult. If a large and high affinity small molecule is found it may provide the required stability to produce diffraction capable crystals in the absence of DNA.

Our efforts to find a novel MBD2 small molecule were not as successful as we had originally hoped, but were also not futile. After a large HTS effort the lead biologically active compounds identified bound to DNA. Though they show specificity by changing chromatin at only methylated promoters, DNA binding small molecules are often difficult to develop and move into the clinic due to limiting toxicity (*201, 202*)(*203*). If these compounds are ever used in animals they will likely have to undergo significant improvement and structure optimization.

Three other small molecules were inactive in all of our assays except TR-FRET. KCC-131, KCC-107 and 127 may interact with DNA or

protein at high micromolar concentrations, but they may also interfere with the assay in some undetermined manner. Regardless, more work will be required to ferret out their target if they are going to move forward as a lead compound.

The final active small molecule, identified in the pilot screen, comes from a long and storied history as a therapeutic agent: the suramin derivative NF449. This anionic family is often found through HTS efforts. The most surprising result shown in this thesis is from the isothermal calorimetry data for this compound. NF449 is capable of specific, saturable inhibition of MBD2 by both the TR FRET assay and ITC. This hints its mechanism is not non-specific, like its structure and story in pubchem might suggest. Despite not showing any biological benefit in initial cell line screening we should seriously consider this as compound for what I refer to as “reverse pharmacophore identification;” in that instead of starting with a small compound with weak interaction and trying to build in sensitivity we have a large compound with moderate to weak interaction and need to shrink it down to the portions that interact with protein. The observation that NF449 is littered with carboxyl groups is also striking. The mechanism behind MBD2-MBD’s preference with this moiety would be very interesting to explain. Maybe our crystals in citrate, if somehow coordinated to the protein, will shed some light on this question.

The final experiment chapter of my thesis presents the efforts I have undertaken to identify DNA methylation alterations tracking with prostate cancer aggressiveness. This is very much a work in progress, but has already shown tremendous promise. Based on the work here, and the group's previously published data using the MBD2-MBD as a tool for enriching methylated DNA, I'm extremely confident the MBD-Seq method is accurately identifying methylated regions in the genome. Furthermore, preliminary analysis shows that we have identified candidate genomic regions with DNA methylation biomarkers which are associated with Gleason patterns and tumor stage. Once the sequencing data is properly normalized for inter sample comparisons, additional biomarkers should also be available given the unbiased nature of these data. Of course significant validation and development of a reproducible method for querying these regions will still be required.

I set out five and a half years ago to make a meaningful contribution to the scientific community through careful and diligent research while I earned a Ph.D. Using a baseball analogy, I don't believe I hit a home run with the work I've accomplished but after suffering through a hurricane, two kids and a lot of headache I believe I've met the goal I set out at the start. I hope the legacy I've left behind in this work, in the Nelson and Yegnasubramanian labs, and in my publications will benefit science and patients in the years to come.

References

1. S. Choudhuri, From Waddington's epigenetic landscape to small noncoding RNA: some important milestones in the history of epigenetics research. *Toxicology mechanisms and methods* **21**, 252 (May, 2011).
2. L. Van Speybroeck, From epigenesis to epigenetics: the case of C. H. Waddington. *Annals of the New York Academy of Sciences* **981**, 61 (Dec, 2002).
3. A. H. Ting, K. M. McGarvey, S. B. Baylin, The cancer epigenome--components and functional correlates. *Genes & development* **20**, 3215 (Dec 1, 2006).
4. A. P. Moczek *et al.*, The role of developmental plasticity in evolutionary innovation. *Proceedings. Biological sciences / The Royal Society* **278**, 2705 (Sep 22, 2011).
5. T. H. Scheike, K. K. Holst, J. B. Hjelmborg, Estimating heritability for cause specific mortality based on twin studies. *Lifetime data analysis* **20**, 210 (Apr, 2014).
6. J. B. Hjelmborg *et al.*, The Heritability of Prostate Cancer in the Nordic Twin Study of Cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, (May 8, 2014).
7. I. Locatelli, A. Rosina, P. Lichtenstein, A. I. Yashin, A correlated frailty model with long-term survivors for estimating the heritability of breast cancer. *Statistics in medicine* **26**, 3722 (Sep 10, 2007).
8. I. Locatelli, P. Lichtenstein, A. I. Yashin, The heritability of breast cancer: a Bayesian correlated frailty model applied to Swedish twins data. *Twin research : the official journal of the International Society for Twin Studies* **7**, 182 (Apr, 2004).
9. W. F. Page, M. M. Braun, A. W. Partin, N. Caporaso, P. Walsh, Heredity and prostate cancer: a study of World War II veteran twins. *The Prostate* **33**, 240 (Dec 1, 1997).
10. N. V. Holm, M. Hauge, B. Harvald, Etiologic factors of breast cancer elucidated by a study of unselected twins. *Journal of the National Cancer Institute* **65**, 285 (Aug, 1980).
11. K. Murray, The Occurrence of Epsilon-N-Methyl Lysine in Histones. *Biochemistry* **3**, 10 (Jan, 1964).
12. E. L. Gershey, G. Vidali, V. G. Allfrey, Chemical studies of histone acetylation. The occurrence of epsilon-N-acetyllysine in the f2a1 histone. *The Journal of biological chemistry* **243**, 5018 (Oct 10, 1968).

13. M. G. Ord, L. A. Stocken, Phosphate and thiol groups in histone f3 from rat liver and thymus nuclei. *The Biochemical journal* **102**, 631 (Feb, 1967).
14. U. Bedi, V. K. Mishra, D. Wasilewski, C. Scheel, S. A. Johnsen, Epigenetic plasticity: A central regulator of epithelial-to-mesenchymal transition in cancer. *Oncotarget* **5**, 2016 (Apr 30, 2014).
15. G. Zhang, S. Pradhan, Mammalian epigenetic mechanisms. *IUBMB life* **66**, 240 (Apr, 2014).
16. A. T. Annunziato, Assembling chromatin: the long and winding road. *Biochimica et biophysica acta* **1819**, 196 (Mar-Apr, 2013).
17. A. J. Bannister, T. Kouzarides, Regulation of chromatin by histone modifications. *Cell research* **21**, 381 (Mar, 2011).
18. C. Sawan, Z. Herceg, Histone modifications and cancer. *Advances in genetics* **70**, 57 (2010).
19. P. Mummaneni, S. S. Shord, Epigenetics and oncology. *Pharmacotherapy* **34**, 495 (May, 2014).
20. B. S. Mann, J. R. Johnson, M. H. Cohen, R. Justice, R. Pazdur, FDA approval summary: vorinostat for treatment of advanced primary cutaneous T-cell lymphoma. *The oncologist* **12**, 1247 (Oct, 2007).
21. StatBite: FDA oncology drug product approvals in 2009. *Journal of the National Cancer Institute* **102**, 219 (Feb 24, 2010).
22. A. Bird, The essentials of DNA methylation. *Cell* **70**, 5 (Jul 10, 1992).
23. K. D. Robertson, P. A. Jones, DNA methylation: past, present and future directions. *Carcinogenesis* **21**, 461 (Mar, 2000).
24. A. M. Deaton, A. Bird, CpG islands and the regulation of transcription. *Genes & development* **25**, 1010 (May 15, 2011).
25. D. M. Messerschmidt, B. B. Knowles, D. Solter, DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes & development* **28**, 812 (Apr 15, 2014).
26. J. M. Sun, V. A. Spencer, H. Y. Chen, L. Li, J. R. Davie, Measurement of histone acetyltransferase and histone deacetylase activities and kinetics of histone acetylation. *Methods* **31**, 12 (Sep, 2003).
27. S. Kangaspeska *et al.*, Transient cyclical methylation of promoter DNA. *Nature* **452**, 112 (Mar 6, 2008).
28. M. Okano, D. W. Bell, D. A. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247 (Oct 29, 1999).
29. T. P. Jurkowski, R. Shanmugam, M. Helm, A. Jeltsch, Mapping the tRNA binding site on the surface of human DNMT2 methyltransferase. *Biochemistry* **51**, 4438 (Jun 5, 2012).

30. P. O. Esteve *et al.*, Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes & development* **20**, 3089 (Nov 15, 2006).
31. J. Song, O. Rechkoblit, T. H. Bestor, D. J. Patel, Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**, 1036 (Feb 25, 2011).
32. J. Song, M. Teplova, S. Ishibe-Murakami, D. J. Patel, Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science* **335**, 709 (Feb 10, 2012).
33. E. Li, T. H. Bestor, R. Jaenisch, Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915 (Jun 12, 1992).
34. J. E. Dodge, B. H. Ramsahoye, Z. G. Wo, M. Okano, E. Li, De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation. *Gene* **289**, 41 (May 1, 2002).
35. D. Meilinger *et al.*, Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO reports* **10**, 1259 (Nov, 2009).
36. X. Cheng, R. M. Blumenthal, Mammalian DNA methyltransferases: a structural perspective. *Structure* **16**, 341 (Mar, 2008).
37. T. Chen, Y. Ueda, S. Xie, E. Li, A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. *The Journal of biological chemistry* **277**, 38746 (Oct 11, 2002).
38. S. K. Ooi *et al.*, DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714 (Aug 9, 2007).
39. D. Jia, R. Z. Jurkowska, X. Zhang, A. Jeltsch, X. Cheng, Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449**, 248 (Sep 13, 2007).
40. K. Hata, M. Okano, H. Lei, E. Li, Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129**, 1983 (Apr, 2002).
41. J. A. Law, S. E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature reviews. Genetics* **11**, 204 (Mar, 2010).
42. R. Z. Jurkowska *et al.*, Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. *Nucleic acids research* **36**, 6656 (Dec, 2008).
43. H. Gowher, K. Liebert, A. Hermann, G. Xu, A. Jeltsch, Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *The Journal of biological chemistry* **280**, 13341 (Apr 8, 2005).
44. C. Holz-Schietinger, N. O. Reich, The inherent processivity of the human de novo methyltransferase 3A (DNMT3A) is enhanced by

DNMT3L. *The Journal of biological chemistry* **285**, 29091 (Sep 17, 2010).
45. D. Bourc'his, G. L. Xu, C. S. Lin, B. Bollman, T. H. Bestor, Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536 (Dec 21, 2001).
46. K. E. Webster *et al.*, Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 4068 (Mar 15, 2005).
47. R. Jaenisch, A. Bird, Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics* **33 Suppl**, 245 (Mar, 2003).
48. S. Laget *et al.*, The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PloS one* **5**, e11982 (2010).
49. R. R. Meehan, J. D. Lewis, S. McKay, E. L. Kleiner, A. P. Bird, Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**, 499 (Aug 11, 1989).
50. J. D. Lewis *et al.*, Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**, 905 (Jun 12, 1992).
51. B. Hendrich, A. Bird, Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and cellular biology* **18**, 6538 (Nov, 1998).
52. X. Nan *et al.*, Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386 (May 28, 1998).
53. N. L. Adkins, P. T. Georgel, MeCP2: structure and function. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **89**, 1 (Feb, 2011).
54. P. A. Jones, D. Takai, The role of DNA methylation in mammalian epigenetics. *Science* **293**, 1068 (Aug 10, 2001).
55. L. S. Weaving, C. J. Ellaway, J. Gecz, J. Christodoulou, Rett syndrome: clinical review and genetic update. *Journal of medical genetics* **42**, 1 (Jan, 2005).
56. J. Guy, B. Hendrich, M. Holmes, J. E. Martin, A. Bird, A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nature genetics* **27**, 322 (Mar, 2001).
57. S. H. Cross, R. R. Meehan, X. Nan, A. Bird, A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nature genetics* **16**, 256 (Jul, 1997).
58. N. Fujita *et al.*, Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA

methylation-based transcriptional repression. *The Journal of biological chemistry* **278**, 24132 (Jun 27, 2003).
59. I. Ohki *et al.*, Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* **105**, 487 (May 18, 2001).
60. H. F. Jorgensen, I. Ben-Porath, A. P. Bird, Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Molecular and cellular biology* **24**, 3387 (Apr, 2004).
61. J. Berger, A. Bird, Role of MBD2 in gene regulation and tumorigenesis. *Biochemical Society transactions* **33**, 1537 (Dec, 2005).
62. Y. Cai *et al.*, The NuRD complex cooperates with DNMTs to maintain silencing of key colorectal tumor suppressor genes. *Oncogene* **33**, 2157 (Apr 24, 2014).
63. S. Yegnasubramanian, X. Lin, M. C. Haffner, A. M. DeMarzo, W. G. Nelson, Combination of methylated-DNA precipitation and methylation-sensitive restriction enzymes (COMPARE-MS) for the rapid, sensitive and quantitative detection of DNA methylation. *Nucleic acids research* **34**, e19 (2006).
64. J. C. Hansen *et al.*, DNA binding restricts the intrinsic conformational flexibility of methyl CpG binding protein 2 (MeCP2). *The Journal of biological chemistry* **286**, 18938 (May 27, 2011).
65. J. Berger, O. Sansom, A. Clarke, A. Bird, MBD2 is required for correct spatial gene expression in the gut. *Molecular and cellular biology* **27**, 4049 (Jun, 2007).
66. B. Hendrich, J. Guy, B. Ramsahoye, V. A. Wilson, A. Bird, Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes & development* **15**, 710 (Mar 15, 2001).
67. O. J. Sansom *et al.*, Deficiency of Mbd2 suppresses intestinal tumorigenesis. *Nature genetics* **34**, 145 (Jun, 2003).
68. X. Le Guezennec *et al.*, MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Molecular and cellular biology* **26**, 843 (Feb, 2006).
69. X. Nan, P. Tate, E. Li, A. Bird, DNA methylation specifies chromosomal localization of MeCP2. *Molecular and cellular biology* **16**, 414 (Jan, 1996).
70. B. Hendrich *et al.*, Genomic structure and chromosomal mapping of the murine and human Mbd1, Mbd2, Mbd3, and Mbd4 genes. *Mammalian genome : official journal of the International Mammalian Genome Society* **10**, 906 (Sep, 1999).
71. M. F. Fraga *et al.*, The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic acids research* **31**, 1765 (Mar 15, 2003).

72. K. L. Ho *et al.*, MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Molecular cell* **29**, 525 (Feb 29, 2008).
73. T. Lang, C. de Chastellier, Fluid phase and mannose receptor-mediated uptake of horseradish peroxidase in mouse bone marrow-derived macrophages. Biochemical and ultrastructural study. *Biology of the cell / under the auspices of the European Cell Biology Organization* **53**, 149 (1985).
74. J. M. Cramer *et al.*, Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA. *The Journal of biological chemistry* **289**, 1294 (Jan 17, 2014).
75. E. Wong *et al.*, Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 14937 (Nov 12, 2002).
76. J. C. Shen, W. M. Rideout, 3rd, P. A. Jones, The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic acids research* **22**, 972 (Mar 25, 1994).
77. B. Hendrich, U. Hardeland, H. H. Ng, J. Jiricny, A. Bird, The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* **401**, 301 (Sep 16, 1999).
78. P. Wu *et al.*, Mismatch repair in methylated DNA. Structure and activity of the mismatch-specific thymine glycosylase domain of methyl-CpG-binding protein MBD4. *The Journal of biological chemistry* **278**, 5285 (Feb 14, 2003).
79. C. B. Millar *et al.*, Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* **297**, 403 (Jul 19, 2002).
80. H. N. Cukier *et al.*, The expanding role of MBD genes in autism: identification of a MECP2 duplication and novel alterations in MBD5, MBD6, and SETDB1. *Autism research : official journal of the International Society for Autism Research* **5**, 385 (Dec, 2012).
81. S. R. Williams *et al.*, Haploinsufficiency of MBD5 associated with a syndrome involving microcephaly, intellectual disabilities, severe speech impairment, and seizures. *European journal of human genetics : EJHG* **18**, 436 (Apr, 2010).
82. C. Bronner, G. Fuhrmann, F. L. Chedin, M. Macaluso, S. Dhe-Paganon, UHRF1 Links the Histone code and DNA Methylation to ensure Faithful Epigenetic Memory Inheritance. *Genetics & epigenetics* **2009**, 29 (Jan 14, 2010).
83. G. V. Avvakumov *et al.*, Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* **455**, 822 (Oct 9, 2008).
84. A. Nishiyama *et al.*, Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature* **502**, 249 (Oct 10, 2013).

85. E. Citterio *et al.*, Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Molecular and cellular biology* **24**, 2526 (Mar, 2004).
86. B. Delatte, R. Deplus, F. Fuks, Playing TETris with DNA modifications. *The EMBO journal* **33**, 1198 (Jun 2, 2014).
87. S. K. Ooi, T. H. Bestor, The colorful history of active DNA demethylation. *Cell* **133**, 1145 (Jun 27, 2008).
88. R. R. Meehan, J. D. Lewis, A. P. Bird, Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic acids research* **20**, 5085 (Oct 11, 1992).
89. H. F. Jorgensen, K. Adie, P. Chaubert, A. P. Bird, Engineering a high-affinity methyl-CpG-binding protein. *Nucleic acids research* **34**, e96 (2006).
90. T. Baubec, R. Ivanek, F. Lienert, D. Schubeler, Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153**, 480 (Apr 11, 2013).
91. R. J. Klose *et al.*, DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Molecular cell* **19**, 667 (Sep 2, 2005).
92. F. Bedogni *et al.*, Rett syndrome and the urge of novel approaches to study MeCP2 functions and mechanisms of action. *Neuroscience and biobehavioral reviews*, (Mar 2, 2014).
93. T. Clouaire, J. I. de Las Heras, C. Merusi, I. Stancheva, Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. *Nucleic acids research* **38**, 4620 (Aug, 2010).
94. C. Mayer-Jung, D. Moras, Y. Timsit, Effect of cytosine methylation on DNA-DNA recognition at CpG steps. *Journal of molecular biology* **270**, 328 (Jul 18, 1997).
95. S. P. Chandler, D. Guschin, N. Landsberger, A. P. Wolffe, The methyl-CpG binding transcriptional repressor MeCP2 stably associates with nucleosomal DNA. *Biochemistry* **38**, 7008 (Jun 1, 1999).
96. R. I. Wakefield *et al.*, The solution structure of the domain from MeCP2 that binds to methylated DNA. *Journal of molecular biology* **291**, 1055 (Sep 3, 1999).
97. I. Ohki, N. Shimotake, N. Fujita, M. Nakao, M. Shirakawa, Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *The EMBO journal* **18**, 6653 (Dec 1, 1999).
98. J. Otani *et al.*, Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. *The Journal of biological chemistry* **288**, 6351 (Mar 1, 2013).

99. J. N. Scarsdale, H. D. Webb, G. D. Ginder, D. C. Williams, Jr., Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic acids research* **39**, 6741 (Aug, 2011).
100. N. Dephoure *et al.*, A quantitative atlas of mitotic phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 10762 (Aug 5, 2008).
101. P. V. Hornbeck *et al.*, PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic acids research* **40**, D261 (Jan, 2012).
102. C. P. Tan, S. Nakielny, Control of the DNA methylation system component MBD2 by protein arginine methylation. *Molecular and cellular biology* **26**, 7224 (Oct, 2006).
103. K. Phillips, A. H. de la Pena, The combined use of the Thermofluor assay and ThermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **Chapter 10**, Unit10 28 (Apr, 2011).
104. C. Broennimann *et al.*, The PILATUS 1M detector. *Journal of synchrotron radiation* **13**, 120 (Mar, 2006).
105. A. J. McCoy *et al.*, Phaser crystallographic software. *Journal of applied crystallography* **40**, 658 (Aug 1, 2007).
106. G. M. Sheldrick, Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta crystallographica. Section D, Biological crystallography* **66**, 479 (Apr, 2010).
107. T. C. Terwilliger, J. Berendzen, Automated MAD and MIR structure solution. *Acta crystallographica. Section D, Biological crystallography* **55**, 849 (Apr, 1999).
108. T. C. Terwilliger, Maximum-likelihood density modification. *Acta crystallographica. Section D, Biological crystallography* **56**, 965 (Aug, 2000).
109. W. Kabsch, Xds. *Acta crystallographica. Section D, Biological crystallography* **66**, 125 (Feb, 2010).
110. N. Wyhs, D. Walker, H. Giovinazzo, S. Yegnasubramanian, W. G. Nelson, Time-Resolved Fluorescence Resonance Energy Transfer Assay for Discovery of Small-Molecule Inhibitors of Methyl-CpG Binding Domain Protein 2. *Journal of biomolecular screening*, (Mar 7, 2014).
111. S. B. Baylin, P. A. Jones, A decade of exploring the cancer epigenome - biological and translational implications. *Nature reviews* **11**, 726 (Oct, 2011).
112. A. Y. Lai, P. A. Wade, Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nature reviews* **11**, 588 (Aug).

113. E. E. Cameron, K. E. Bachman, S. Myohanen, J. G. Herman, S. B. Baylin, Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nat Genet* **21**, 103 (Jan, 1999).
114. J. M. Foulks *et al.*, Epigenetic drug discovery: targeting DNA methyltransferases. *J Biomol Screen* **17**, 2 (Jan).
115. A. Mahindra *et al.*, Latest advances and current challenges in the treatment of multiple myeloma. *Nat Rev Clin Oncol* **9**, 135 (Mar).
116. M. Rius, F. Lyko, Epigenetic cancer therapy: rationales, targets and drugs. *Oncogene* **31**, 4257 (Sep 27).
117. X. Lin, W. G. Nelson, Methyl-CpG-binding domain protein-2 mediates transcriptional repression associated with hypermethylated GSTP1 CpG islands in MCF-7 breast cancer cells. *Cancer Res* **63**, 498 (Jan 15, 2003).
118. L. K. Su *et al.*, Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* **256**, 668 (May 1, 1992).
119. H. Bazin, E. Trinquet, G. Mathis, Time resolved amplification of cryptate emission: a versatile technology to trace biomolecular interactions. *J Biotechnol* **82**, 233 (Jan, 2002).
120. S. M. van den Wildenberg, B. Prevo, E. J. Peterman, A brief introduction to single-molecule fluorescence methods. *Methods in molecular biology (Clifton, N.J* **783**, 81 (2011).
121. K. M. Dawson, Activity of SC33428, a novel bishydrazone-bridged derivative of 4-demethoxydaunorubicin, against experimental tumors in mice. *Cancer Res* **43**, 2880 (Jun, 1983).
122. T. D. Shenkenberg, D. D. Von Hoff, Mitoxantrone: a new anticancer drug with significant clinical activity. *Annals of internal medicine* **105**, 67 (Jul, 1986).
123. T. Blumenthal, T. A. Landers, The inhibition of nucleic acid-binding proteins by aurintricarboxylic acid. *Biochemical and biophysical research communications* **55**, 680 (Dec 10, 1973).
124. R. Hausmann, G. Schmalzing, P2X1 and P2X2 receptors in the central nervous system as possible drug targets. *CNS & neurological disorders drug targets* **11**, 675 (Sep, 2012).
125. S. Oka *et al.*, NMR structure of transcription factor Sp1 DNA binding domain. *Biochemistry* **43**, 16027 (Dec 28, 2004).
126. M. Hulsmann *et al.*, NF449, a novel picomolar potency antagonist at human P2X1 receptors. *European journal of pharmacology* **470**, 1 (May 30, 2003).
127. S. El-Ajouz, D. Ray, R. C. Allsopp, R. J. Evans, Molecular basis of selective antagonism of the P2X1 receptor for ATP by NF449 and suramin: contribution of basic amino acids in the cysteine-rich loop. *British journal of pharmacology* **165**, 390 (Jan, 2012).

128. R. A. North, Molecular physiology of P2X receptors. *Physiological reviews* **82**, 1013 (Oct, 2002).
129. G. Burnstock, Pathophysiology and therapeutic potential of purinergic signaling. *Pharmacological reviews* **58**, 58 (Mar, 2006).
130. M. Hohenegger *et al.*, Gsalpha-selective G protein antagonists. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 346 (Jan 6, 1998).
131. P. Krejci *et al.*, NF449 is a novel inhibitor of fibroblast growth factor receptor 3 (FGFR3) signaling active in chondrocytes and multiple myeloma cells. *The Journal of biological chemistry* **285**, 20644 (Jul 2, 2010).
132. R. D. Team, R: A Language and Environment For Statistical Computing. *R Foundation For Statistical Computing* **{ISBN} 3-900051-07-0**, (2011).
133. J. H. Zhang, T. D. Chung, K. R. Oldenburg, A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen* **4**, 67 (1999).
134. S. Yegnasubramanian *et al.*, DNA hypomethylation arises later in prostate cancer progression than CpG island hypermethylation and contributes to metastatic tumor heterogeneity. *Cancer research* **68**, 8954 (Nov 1, 2008).
135. M. J. Aryee *et al.*, DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. *Science translational medicine* **5**, 169ra10 (Jan 23, 2013).
136. A. P. Feinberg *et al.*, Personalized epigenomic signatures that are stable over time and covary with body mass index. *Science translational medicine* **2**, 49ra67 (Sep 15, 2010).
137. K. E. Schuebel *et al.*, Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. *PLoS genetics* **3**, 1709 (Sep, 2007).
138. S. R. Floyd *et al.*, The bromodomain protein Brd4 insulates chromatin from DNA damage signalling. *Nature* **498**, 246 (Jun 13, 2013).
139. A. Muller, M. Florek, 5-Azacytidine/Azacitidine. *Recent results in cancer research. Fortschritte der Krebsforschung. Progres dans les recherches sur le cancer* **184**, 159 (2010).
140. M. L. Orta *et al.*, 5-Aza-2'-deoxycytidine causes replication lesions that require Fanconi anemia-dependent homologous recombination for repair. *Nucleic acids research* **41**, 5827 (Jun, 2013).
141. A. Gnyszka, Z. Jastrzebski, S. Flis, DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer. *Anticancer research* **33**, 2989 (Aug, 2013).

142. J. M. Mehnert, W. K. Kelly, Histone deacetylase inhibitors: biology and mechanism of action. *Cancer J* **13**, 23 (Jan-Feb, 2007).
143. C. P. Miller, M. M. Singh, N. Rivera-Del Valle, C. A. Manton, J. Chandra, Therapeutic strategies to enhance the anticancer efficacy of histone deacetylase inhibitors. *Journal of biomedicine & biotechnology* **2011**, 514261 (2011).
144. F. Carrier, Chromatin Modulation by Histone Deacetylase Inhibitors: Impact on Cellular Sensitivity to Ionizing Radiation. *Molecular and cellular pharmacology* **5**, 51 (Jan 1, 2013).
145. J. Yoo, J. H. Kim, K. D. Robertson, J. L. Medina-Franco, Molecular modeling of inhibitors of human DNA methyltransferase with a crystal structure: discovery of a novel DNMT1 inhibitor. *Advances in protein chemistry and structural biology* **87**, 219 (2012).
146. W. S. Xu, R. B. Parmigiani, P. A. Marks, Histone deacetylase inhibitors: molecular mechanisms of action. *Oncogene* **26**, 5541 (Aug 13, 2007).
147. A. Fischer, F. Sananbenesi, A. Mungenast, L. H. Tsai, Targeting the correct HDAC(s) to treat cognitive disorders. *Trends in pharmacological sciences* **31**, 605 (Dec, 2010).
148. M. Yoshida *et al.*, Histone deacetylase as a new target for cancer chemotherapy. *Cancer chemotherapy and pharmacology* **48 Suppl 1**, S20 (Aug, 2001).
149. S. R. Haynes *et al.*, The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. *Nucleic acids research* **20**, 2603 (May 25, 1992).
150. P. Filippakopoulos *et al.*, Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214 (Mar 30, 2012).
151. C. H. Arrowsmith, C. Bountra, P. V. Fish, K. Lee, M. Schapira, Epigenetic protein families: a new frontier for drug discovery. *Nature reviews. Drug discovery* **11**, 384 (May, 2012).
152. M. Rosner, M. Hengstschlager, Targeting epigenetic readers in cancer. *The New England journal of medicine* **367**, 1764 (Nov, 2012).
153. C. Dhalluin *et al.*, Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491 (Jun 3, 1999).
154. C. W. Chung, A. W. Dean, J. M. Woolven, P. Bamborough, Fragment-based discovery of bromodomain inhibitors part 1: inhibitor binding modes and implications for lead discovery. *Journal of medicinal chemistry* **55**, 576 (Jan 26, 2012).
155. P. Bamborough *et al.*, Fragment-based discovery of bromodomain inhibitors part 2: optimization of phenylisoxazole sulfonamides. *Journal of medicinal chemistry* **55**, 587 (Jan 26, 2012).
156. P. Filippakopoulos *et al.*, Selective inhibition of BET bromodomains. *Nature* **468**, 1067 (Dec 23, 2010).

157. E. Nicodeme *et al.*, Suppression of inflammation by a synthetic histone mimic. *Nature* **468**, 1119 (Dec 23, 2010).
158. I. A. Asangani *et al.*, Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature*, (Apr 23, 2014).
159. J. E. Delmore *et al.*, BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904 (Sep 16, 2011).
160. O. Y. Mian *et al.*, Methyl-binding domain protein 2-dependent proliferation and survival of breast cancer cells. *Molecular cancer research : MCR* **9**, 1152 (Aug, 2011).
161. S. N. Krishna *et al.*, A fluorescence-based thermal shift assay identifies inhibitors of mitogen activated protein kinase kinase 4. *PloS one* **8**, e81504 (2013).
162. K. Peltonen *et al.*, A targeting modality for destruction of RNA polymerase I that possesses anticancer activity. *Cancer cell* **25**, 77 (Jan 13, 2014).
163. Y. Chi, K. Gao, H. Zhang, M. Takeda, J. Yao, Suppression of cell membrane permeability by suramin: involvement of its inhibitory actions on connexin 43 hemichannels. *British journal of pharmacology*, (Mar 18, 2014).
164. W. Y. Bolton E, Thiessen PA, Bryant SH, *PubChem: Integrated Platform of Small Molecules and Biological Activities*. Chapter 12 In Annual Reports in Computational Chemistry (American Chemical Society, 2008), vol. 4.
165. R. C. T. (2013). (Vienna, Austria, 2013).
166. M. C. Haffner *et al.*, Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements. *Nature genetics* **42**, 668 (Aug, 2010).
167. W. G. e. a. Nelson. M. D. Abeloff, Ed., Clinical Oncology (Elsevier Churchill Livingstone, Philadelphia, PA, ed. 3rd, 2004).
168. A. S. Donn, C. S. Muir, Prostatic cancer: some epidemiological features. *Bulletin du cancer* **72**, 381 (1985).
169. R. Etzioni *et al.*, Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. *Journal of the National Cancer Institute* **94**, 981 (Jul 3, 2002).
170. R. R. Gatling, Prostate carcinoma: an autopsy evaluation of the influence of age, tumor grade, and therapy on tumor biology. *Southern medical journal* **83**, 782 (Jul, 1990).
171. P. C. Walsh, T. L. DeWeese, M. A. Eisenberger, Clinical practice. Localized prostate cancer. *The New England journal of medicine* **357**, 2696 (Dec 27, 2007).
172. B. A. Morris, J. D. Robinson, E. Piall, G. W. Aherne, V. Marks, Proceedings: Development of a radioimmunoassay for morphine having minimal cross-reactivity with codeine. *The Journal of endocrinology* **64**, 6P (Jan, 1975).

173. D. Hanahan, R. A. Weinberg, The hallmarks of cancer. *Cell* **100**, 57 (Jan 7, 2000).
174. Z. Chen, L. Wang, Q. Wang, W. Li, Histone modifications and chromatin organization in prostate cancer. *Epigenomics* **2**, 551 (Aug, 2010).
175. S. Yegnasubramanian, W. G. Nelson, in *DNA Methylation, Epigenetics and Metastasis,* M. Esteller, Ed. (Springer publishers, The Netherlands, 2005), pp. 45-79.
176. J. G. Herman, S. B. Baylin, Gene silencing in cancer in association with promoter hypermethylation. *The New England journal of medicine* **349**, 2042 (Nov 20, 2003).
177. S. Yegnasubramanian *et al.*, Chromosome-wide mapping of DNA methylation patterns in normal and malignant prostate cells reveals pervasive methylation of gene-associated and conserved intergenic sequences. *BMC genomics* **12**, 313 (2011).
178. D. Serre, B. H. Lee, A. H. Ting, MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic acids research* **38**, 391 (Jan, 2010).
179. S. Yegnasubramanian *et al.*, Hypermethylation of CpG islands in primary and metastatic human prostate cancer. *Cancer research* **64**, 1975 (Mar 15, 2004).
180. J. D. Brooks *et al.*, CG island methylation changes near the GSTP1 gene in prostatic intraepithelial neoplasia. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **7**, 531 (Jun, 1998).
181. M. C. Haffner *et al.*, Tracking the clonal origin of lethal prostate cancer. *The Journal of clinical investigation* **123**, 4918 (Nov 1, 2013).
182. M. J. Aryee *et al.*, Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363 (May 15, 2014).
183. A. R. Rao, H. G. Motiwala, O. M. Karim, The discovery of prostate-specific antigen. *BJU international* **101**, 5 (Jan, 2008).
184. J. B. Eifler *et al.*, An updated prostate cancer staging nomogram (Partin tables) based on cases from 2006 to 2011. *BJU international* **111**, 22 (Jan, 2013).
185. M. Darshan *et al.*, Biobanking of derivatives from radical retropubic and robot-assisted laparoscopic prostatectomy tissues as part of the prostate cancer biorepository network. *The Prostate* **74**, 61 (Jan, 2014).
186. W. Liu *et al.*, Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. *Nature medicine* **15**, 559 (May, 2009).

187. S. Yegnasubramanian, Preparation of fragment libraries for next-generation sequencing on the applied biosystems SOLiD platform. *Methods in enzymology* **529**, 185 (2013).
188. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).
189. J. Feng, T. Liu, Y. Zhang, Using MACS to identify peaks from ChIP-Seq data. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **Chapter 2**, Unit 2 14 (Jun, 2011).
190. J. Feng, T. Liu, B. Qin, Y. Zhang, X. S. Liu, Identifying ChIP-seq enrichment using MACS. *Nature protocols* **7**, 1728 (Sep, 2012).
191. M. Lawrence *et al.*, Software for computing and annotating genomic ranges. *PLoS computational biology* **9**, e1003118 (2013).
192. H. Wickham. (Springer, NY, 2009).
193. D. Adler. (2005).
194. F. J. Harrell, C. Dupont, e. al, F. E. H. Jr, Ed. (2014).
195. M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik. (2014).
196. Gregory R. Warnes *et al.* (2014).
197. J. Zhang, P. L. Yang, N. S. Gray, Targeting cancer with small molecule kinase inhibitors. *Nature reviews. Cancer* **9**, 28 (Jan, 2009).
198. G. Qu *et al.*, Combination of BIBW2992 and ARQ 197 is effective against erlotinib-resistant human lung cancer cells with the EGFR T790M mutation. *Oncology reports* **32**, 341 (Jul, 2014).
199. L. Formisano *et al.*, Epidermal growth factor receptor activation modulates Src-dependent resistance to lapatinib in breast cancer models. *Breast cancer research : BCR* **16**, R45 (May 5, 2014).
200. C. N. Sternberg, D. P. Petrylak, R. A. Madan, C. Parker, Progress in the treatment of advanced prostate cancer. *American Society of Clinical Oncology educational book / ASCO. American Society of Clinical Oncology. Meeting* **34**, 117 (2014).
201. L. H. Hurley, DNA and its associated processes as targets for cancer therapy. *Nature reviews. Cancer* **2**, 188 (Mar, 2002).
202. R. Palchaudhuri, P. J. Hergenrother, DNA as a target for anticancer compounds: methods to determine the mode of binding and the mechanism of action. *Current opinion in biotechnology* **18**, 497 (Dec, 2007).
203. K. Gurova, New hopes from old drugs: revisiting DNA-binding small molecules as anticancer agents. *Future Oncol* **5**, 1685 (Dec, 2009).

Curriculum Vitae

Nicolás Wyhs

9993 Sherwood Farm Rd
Owings Mills, MD 21117
Cell: (617) 610-8696
Email: nawyhs@gmail.com

Education:

Doctor of Philosophy	Johns Hopkins School of Medicine Department of Pharmacology and Molecular Sciences	June 2014
Bachelor of Science (BS)	Massachusetts Institute of Technology Chemical Engineering/Minor in Spanish	June 2005

Academic Research Experience:

Johns Hopkins School of Medicine-Department of Pharmacology — Baltimore, MD
Candidate for Ph.D. — Mentor: Dr. William G Nelson — 8/2008-Present

Graduate student in the lab of Dr. William Nelson and Dr. Srinivasan Yegnasubramanian studying methylated DNA reader proteins.

- Characterized the biophysics of the MBD2 protein's interaction with methylated DNA by solving the x-ray crystallography structure and through biochemical assays such as ITC, fluorescence polarization, thermal melt analysis, and point mutation studies
- Established and maintained collaborations with principle investigators in other fields at our institution to bring the techniques of x-ray crystallography and pathology to our lab
- First to apply Time Resolved Fluorescence Energy Transfer (TR-FRET) Technology to monitor DNA-protein interactions. Utilized this novel technology to assemble a HTS assay to find small molecule inhibitors of our epigenetic reader protein
- Co-authored a funded R03 grant through the NCI to cover expenses for applying our TR FRET HTS assay to their 370,000 compound small molecule library. Collaborated with medicinal chemistry and HTS departments at the Scripps Institute® to complete the screen.
- Validated 2 compounds from the HTS screen as able to relieve epigenetic repression in multiple cancer cell line models. Manuscript describing the screen and validation of these compounds in progress
- Utilized a novel methylated DNA precipitation strategy with next generation sequencing to analyze the prostate cancer methylome for signatures that could improve prognosis during biopsy.
- Authored and awarded a 3-year grant to cover salary and travel related to my work with the prostate cancer methylome analysis
- Served as a peer mentor for the medical school graduate students from 2010-2014, assisting more than 200 students in their transition between undergraduate and graduate school and through various milestones of graduate school in group or one-on-one sessions

- From 2009-2014 I was responsible for managing 20-40 prospective students each year during our department's interview weekend for the first of three days they were visiting Johns Hopkins.
- Mentored 3 junior graduate students and 2 high school students during research rotations

Massachusetts Institute of Technology Cambridge, MA
BS in Chemical Engineering Mentor: Dr. Daniel I.C. Wang 8/2001-6/2005
Received a Bachelor of Science in Chemical Engineering and a minor in Spanish studies in June 2005.

- Chemical Engineering Project Lab- In a small group successfully engineered an incubator capable of hypoxic tissue culture growth under sterile conditions. Applied incubator to determine the optimal growth conditions for mouse embryonic stem cells
- Biological Engineering Laboratory Course- Expressed, purified and confirmed presence of a protein of interest in bacterial and mammalian systems
- Undergraduate Research Opportunities Program (UROP)-Worked in the Biotechnology Process Engineering Center (BPEC) under Dr. Peng Jiao and Dr. Daniel I.C. Wang. Successfully engineered a bacterial line with a modular coding sequence cassette for overexpression of a gene of interest.

Work Experience:

Ambergen Incorporated Watertown, MA
Research Associate Supervisor: Dr. Christopher Sears 12/2007-8/2008

Project leader investigating novel means of determining prognosis in colorectal cancer (CRC) via a gene expression platform.

- Analyzed and transitioned gene expression microarray data into a biomarker panel for prognosis
- Designed and implemented a reproducible RNA extraction protocol from paraffin tissue after overcoming numerous technical issues related to nucleic acid damage from fixing process and age (patent pending)
- Successfully created a real time PCR based panel for the biomarkers discovered in microarray analysis and applied it to a large validation set of RNA extracted from FFPE tissue
- Worked with principle investigator and CSO to guide overall project design in meetings and via written progress reports

Dana Farber Cancer Institute-Harvard Medical School Boston, MA
Research Technician Supervisor: Dr. Levi Garraway & Matthew Meyerson Full Time 9/2006-12/2007

Technician in a lab focused on the genomic understanding of melanoma, lung and prostate cancer.

- Independently designed and executed experiments related to confirming whole genome sequencing study results from senior lab members using Sequenome® technology
- Helped expand a mass spectrometry based high throughput assay (Oncomap) from covering a few hundred to over a thousand known point mutations and applied it to hundreds of patient samples

MIT Nuclear Research Reactor Lab — Cambridge, MA
Senior Reactor Operator/Supervisor — Supervisor: Edward Lau — 4/2002-8/2008

Nuclear Regulatory Commission (NRC) licensed reactor supervisor for this 5 megawatt research reactor.

- NRC Licensed Reactor Operator 2002-2004.
- NRC Licensed Senior Reactor Operator-Supervisor 2004-2008.
- Responsible for all aspects of the 5MW nuclear reactor while on duty including supervision and calculation of startup conditions, monitoring systems for abnormal indications and radiation levels.
- Maintained ultimate responsibility for deciding appropriate responses to problems/alarms
- Maintained expert knowledge of all emergency protocols and procedures in case of catastrophic disaster. Had this occurred the supervisor on duty acts as the initial incident commander
- Interacted regularly with other public entities (fire, EMS and police department staff) to prepare for incidents
- Supervised >20 reactor operators to ensure safe operation of the reactor
- Maintained extensive knowledge of radiation safety and methods for working in high radiation fields
- Regularly assisted lab engineers with routine maintenance of mechanical and electronic systems as well as refueling efforts

Clinical Work Experience:

Bel Air Volunteer Fire Company — Bel Air, MD
Paramedic — Supervisor: Loren Anderson — 5/2011-Present
Work as a paid employee for this volunteer service that covers 911 emergencies in the town of Bel Air in Harford County Maryland.

Liberty Road Volunteer Fire company — Randallstown, MD
Paramedic — Supervisor: James Fulton — 9/2010-2/2014

Cataldo Ambulance Service Inc — Somerville, MA
Paramedic — Supervisor: Human Resources Department — 6/2002-8/2008
Served in a high public visibility position as ice/court paramedic for two professional sports teams (Boston Celtics and Boston Bruins). This required providing medical care and transport to players, coaches and league officials while also managing the attention of the national media

Community Service:

Owings Mills Volunteer Fire Company — Owings Mills, MD — 4/2009-Present
Perform the duties of a Paramedic as a volunteer for the fire company's ambulance

Volunteer EMT-Basic and Chief of MIT-EMS — Cambridge, MA — 9/2001-8/2008
Co-founder and previous chief of the MIT Ambulance Service. Was responsible for overseeing all aspects of this basic life support transporting ambulance including operations, clinical care and training. Managed between 40-60 EMT providers. Served

as lead contact for MIT administration and other city emergency services (fire, police, etc)

New Milford Community Ambulance Corporation, New Milford, CT 6/2001-6/2003

Professional Societies:
Society of Hispanic Professional Engineers Member/Treasurer 2003-2005
Served as treasurer in 2004. Authored quarterly budget requests for funding from the Institute and managed all assets of approximately $2000-$5000. Responsible for approving all expenditures and reimbursements. Invested significant time as a volunteer in the Jamaica Plains area of Boston working with more than 10 underprivileged children assisting them with their studies and preparation for the SAT.

Society for Advancement of Chicanos and Native Americans in Science
Member 2010-Present
Selected as the student representative of the Johns Hopkins School of Medicine at this national conference.

Publications:
Wyhs N, Walker D, Yegnasubramanian S, Nelson WG. Time Resolved Fluorescence Resonance Energy Transfer assay for Discovery of Small molecule Inhibitors of Methyl-CpG Binding Domain Protein 2.
J Biomol Screen. 2014 Mar 7. [Epub ahead of print].

Darshan M, Zheng Q, Fedor HL, **Wyhs N**, Yegnasubramanian S, Lee P, Melamed J, Netto GJ, Trock BJ, De Marzo AM, Sfanos KS. Biobanking of derivatives from radical retropubic and robot-assisted laparoscopic prostatectomy tissues as part of the prostate cancer biorepository network. Prostate. 2013 Sep 21.

Dutt A, Salvesen HB, Chen TH, Ramos AH, Onofrio RC, Hatton C, Nicoletti R, Winckler W, Grewal R, Hanna M, **Wyhs N**, Ziaugra L, Richter DJ, Trovik J, Engelsen IB, Stefansson IM, Fennell T, Cibulskis K, Zody MC, Akslen LA, Gabriel S, Wong KK, Sellers WR, Meyerson M, Greulich H. Drug-sensitive FGFR2 mutations in endometrial carcinoma. Proc Natl Acad Sci U S A. 2008 Jun 24;105(25):8713-7

Invited Presentations:
Wyhs N, Walker D, Yegnasubramanian S, Nelson WG. Development of a High Throughput Small Molecule Discovery Assay for the Epigenetic Target Methyl-CpG Binding Domain 2. Invited Presentation at: Prostate Specialized Programs of Research Excellence. 5th Annual Conference; 2012 Mar; Fort Lauderdale, FL

Posters:
Wyhs N, Giovinazzo H, Walker D, Yegnasubramanian S, Nelson WG. Identification of Novel Small Molecule Inhibitors of the Epigenetic Reader Protein MBD2. Poster Presentation: AACR 105th Annual Conference; 2014 Apr; San Diego, CA.

Wyhs N, Reichert Z, Yegnasubramanian S, Nelson WG. Biophysical Characterization of Methyl-CpG Binding Domain 2 Protein. Poster Presented At: Prostate Specialized Programs of Research Excellence. 3rd Annual Conference; 2010 Apr; Fort Lauderdale, FL.

Honors/Awards:

Department of Defense Prostate Cancer Research Program Prostate Cancer Pre-Doctoral Training Award 8/2011-8/2014
Authored and awarded grant worth approximately $100,000 covering my full stipend, health insurance and travel expenses.

Undergraduate Leadership Award 5/2005
Awarded by the Public Service Center to one undergraduate representing inspiration and dedication towards community service.

President's Call to Service Lifetime Achievement 12/2005
Awarded by MIT Emergency Medical Service and the President's Council for volunteering over 4000 hours.

Citizen Recognition Award 6/2004
Highest citizen medal awarded by MIT Police Department for outstanding work for the MIT community and public safety.

William L. Stewart, Jr. Award 5/2003
Recognized for outstanding contributions by an individual student or student organization to extracurricular activities.

Multi-Cultural Community Service Award 5/2003
Awarded by the office of Minority Education for great contributions to the community.

Skills and Other Certifications:

Proficient in Spanish
Excellent knowledge of excel, word, power point, and basic computer skills. Experienced with matlab and R programming languages.
Hobbies include scuba diving, racquetball, tennis and hiking.